

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 528 963**

51 Int. Cl.:

C40B 40/10

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.04.2009** **E 09735791 (7)**

97 Fecha y número de publicación de la concesión europea: **22.10.2014** **EP 2281078**

54 Título: **Bibliotecas de paquetes genéticos que comprenden nuevos diseños de CDR1, CDR2, y CDR3 de HC y nuevos diseños de CDR1, CDR2, y CDR3 de LC**

30 Prioridad:

24.04.2008 US 47529 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
13.02.2015

73 Titular/es:

DYAX CORP. (100.0%)
55 Network Drive
Burlington, MA 01803, US

72 Inventor/es:

LADNER, ROBERT C.

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 528 963 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Bibliotecas de paquetes genéticos que comprenden nuevos diseños de CDR1, CDR2, y CDR3 de HC y nuevos diseños de CDR1, CDR2, y CDR3 de LC

REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

- 5 Esta solicitud reivindica la prioridad a la Solicitud U.S. Serie nº 61/047.529, presentada el 24 de abril de 2008. La descripción de la solicitud previa se considera parte de la descripción de esta solicitud.

ANTECEDENTES

- Actualmente es práctica corriente en la técnica preparar bibliotecas de paquetes genéticos que presentan individualmente, presentan y expresan, o comprenden un miembro de una familia diversa de péptidos, polipéptidos o proteínas, y en conjunto presentan, presentan y expresan, o comprenden al menos una porción de la diversidad de aminoácidos de la familia. En muchas bibliotecas corrientes, los péptidos, polipéptidos o proteínas están relacionados con anticuerpos (por ejemplo, Fv monocatenario (scFv), Fv, Fab, anticuerpos o minicuerpos completos (es decir, dímeros constituidos por V_H unida a V_L)). A menudo, comprenden una o más de las CDRs y regiones del marco de las cadenas pesada y ligera de los anticuerpos humanos.
- 15 Las bibliotecas de péptidos, polipéptidos o proteínas se han producido de muchas maneras. Véase, por ejemplo, Knappik et al., J. Mol. Biol., 296, págs. 57-86, (2000), que se incorpora a la presente memoria como referencia. Un método consiste en capturar la diversidad de donantes nativos, ya sea sin tratamiento previo o inmunizados. Otra manera consiste en generar bibliotecas con diversidad sintética. Un tercer método es una combinación de los dos primeros. Por lo general, la diversidad producida por estos métodos se limita a la diversidad de secuencias, es decir, cada miembro de la biblioteca tiene la misma longitud pero difiere de los otros miembros de la familia porque tiene diferentes aminoácidos o variegación en una posición dada en la cadena peptídica, polipeptídica o proteica. Los péptidos, polipéptidos o proteínas naturalmente diversos, sin embargo, no se limitan a la diversidad solamente en sus secuencias de aminoácidos. Por ejemplo, los anticuerpos humanos no se limitan a la diversidad de secuencias en sus aminoácidos; también son diversos en las longitudes de sus cadenas de aminoácidos.
- 20 Hoet et al., Nature Biotechnology, Vol. 23 (3), 1 de marzo de 2005, páginas 344-348, describe la generación de anticuerpos humanos de alta afinidad combinando una diversidad de regiones determinantes de la complementariedad derivadas de donantes y sintéticas.
- Para anticuerpos, la diversidad en longitud de la cadena pesada se produce, por ejemplo, durante las reorganizaciones de la región variable. Véase, por ejemplo, Corbett et al., J. Mol. Biol., 270, págs. 587-97 (1997). La unión de los genes V con los genes J, por ejemplo, da como resultado la inclusión de un segmento D reconocible en CDR3 en alrededor de la mitad de las secuencias de anticuerpos de cadena pesada, creando de este modo regiones que codifican longitudes variables de aminoácidos. Los segmentos D son más habituales en anticuerpos que tienen CDR3s de HC largas. Lo siguiente también puede producirse durante la unión de los segmentos génicos del anticuerpo: (i) el extremo del gen V puede tener cero o varias bases suprimidas o cambiadas; (ii) el extremo del segmento D puede tener cero a muchas bases eliminadas o cambiadas; (iii) numerosas bases aproximadamente al azar pueden insertarse entre V y D o entre D y J; y (iv) se puede editar el extremo 5' de J para eliminar o cambiar varias bases. Estas reorganizaciones dan como resultado anticuerpos que son diversos tanto en la secuencia de aminoácidos como en longitud. Las CDR3s de HC de diferentes longitudes se pueden plegar en diferentes formas, dando las nuevas formas de anticuerpos con los que se unirán los antígenos. Las conformaciones dependen tanto de la longitud como de la secuencia de la CDR3. Se debería recordar que una CDR3 de HC de longitud 8, por ejemplo, y de cualquier secuencia no puede imitar adecuadamente el comportamiento de una CDR3 de longitud 22, por ejemplo.
- Las bibliotecas que contienen solamente diversidad de secuencia de aminoácidos están, de este modo, en desventaja por cuanto no reflejan la diversidad natural del péptido, polipéptido o proteína que la biblioteca pretende imitar. Además, la diversidad en longitud puede ser importante para el funcionamiento en última instancia de la proteína, péptido o polipéptido. Por ejemplo, con respecto a una biblioteca que comprende regiones de anticuerpos, muchos de los péptidos, polipéptidos, proteínas presentados, presentados y expresados, o comprendidos por paquetes genéticos de la biblioteca no pueden plegarse de manera apropiada, o su unión a un antígeno puede resultar desventajosa, si la diversidad tanto en la secuencia como en la longitud no están representadas en la biblioteca.
- Otro inconveniente adicional de tales bibliotecas de paquetes genéticos que presentan, presentan y expresan, o comprenden péptidos, polipéptidos y proteínas es que no están focalizadas en aquellos miembros que están basados en la diversidad de origen natural y de este modo en los miembros que es más probable que sean funcionales y menos probable que sean inmunogénicos. Más bien, las bibliotecas, típicamente, intentan incluir tanta diversidad o variegación en cada posición de CDR como sea posible. Esto hace que la construcción de la biblioteca emplee mucho tiempo y sea menos eficiente de lo necesario. El gran número de miembros que se producen tratando de capturar la diversidad completa también hace la identificación más problemática de lo que sería

necesario. Esto es particularmente cierto dado que muchos miembros de la biblioteca no serán funcionales y serán pegajosos de forma no específica.

Además de la labor de construir bibliotecas sintéticas está la cuestión de la inmunogenicidad. Por ejemplo, hay bibliotecas en las que todos los restos de CDR son Tyr (Y) o Ser (S). Aunque los anticuerpos (Abs) seleccionados de estas bibliotecas muestran afinidad y especificidad elevadas, su composición muy poco habitual los hacen inmunogénicos. La presente invención se refiere a la obtención de Abs que bien podrían proceder del sistema inmunitario humano y de este modo son probablemente menos inmunogénicos. Las bibliotecas de la presente invención retienen tantos restos procedentes de fusiones V-D-J o V-J como sea posible. Para reducir el riesgo de inmunogenicidad, puede ser prudente cambiar cada aminoácido que no sea de la línea germinal tanto en el marco como en las CDRs nuevamente a la línea germinal para determinar si el cambio desde la línea germinal es necesario para retener la afinidad de unión. De este modo, una biblioteca que está desviada en cada posición variada con respecto a la línea germinal reducirá la probabilidad de aislar Abs que tienen aminoácidos no necesarios que no son de la línea germinal.

Los Abs son proteínas grandes y están sujetos a diversas formas de degradación. Una forma de degradación es la desamidación de los restos Asn y Gln (especialmente en Asn-Gly o Gln-Gly) y la isomerización de restos de Asp. Otra forma de degradación es la oxidación de metioninas, triptófano, y cisteína. Otra forma de degradación es la escisión de dipéptidos Asp-Pro. Otra forma de degradación es la formación de piroglutamato a partir de Glu o Gln N-terminal. Es ventajoso proporcionar una biblioteca en la que se minimice la aparición de secuencias problemáticas.

Se proporcionan bibliotecas de vectores o paquetes que codifican miembros de una familia diversa de anticuerpos humanos que comprenden CDR3s de cadena pesada (HC) que tienen una longitud entre alrededor de 3 aminoácidos y alrededor de 35 aminoácidos. Las CDR3s de HC también pueden ser, en ciertas realizaciones, ricas en Tyr (Y) y Ser (S), y/o comprometer regiones D diversificadas y/o comprometer regiones JH extendidas. Por ejemplo, las CDR3s de HC pueden contener más de alrededor de 40% (por ejemplo, entre alrededor de 43% y alrededor de 80%; por ejemplo, más de alrededor de 40% pero menos de alrededor de 100%) de restos Y y/o S, por ejemplo, como se proporciona en los ejemplos aquí. También se proporcionan bibliotecas focalizadas que comprenden tales CDR3s de HC. También se proporcionan diseños para CDR1 de HC, CDR2 de HC, y una biblioteca de VKIII A27 con diversidad en las CDRs. Una biblioteca de vectores o paquetes que codifica miembros de una familia diversa de anticuerpos humanos que comprenden CDR3s de HC descritos aquí puede tener además diversidad en una o más (por ejemplo, en una, dos o tres) de CDR1 de HC, CDR2 de HC, CDR1 de LC, CDR2 de LC, y CDR3 de LC. Por ejemplo, la biblioteca puede tener diversidad en una o más (por ejemplo, en una, dos o tres) de CDR1 de HC, CDR2 de HC, CDR1 de LC, CDR2 de LC, y CDR3 de LC como se describe aquí.

Una región D diversificada es una región D en la que se han introducido uno o más cambios de aminoácidos (por ejemplo, en comparación con la secuencia de una región D de origen natural; por ejemplo, un codón de parada se puede cambiar a un resto Tyr).

Una región JH extendida es una región JH que tiene uno o más restos de aminoácidos presentes en el término amino de la secuencia del marco de la región JH (por ejemplo, aminoterminal a las secuencias de FR4, por ejemplo, que comienzan con WGQ ...). Por ejemplo, JH1 es una región JH extendida. Como otros ejemplos JH2, JH3, JH4, JH5, y JH6 son regiones JH extendidas.

También se proporcionan métodos para obtener y cribar las bibliotecas anteriores y las CDR3s de HC y los anticuerpos obtenidos en tal cribado. También se describen aquí composiciones y kits para la práctica de estos métodos.

En algunos aspectos, la descripción se refiere a una biblioteca focalizada de vectores o paquetes genéticos que presentan, presentan y expresan, o comprenden un miembro de una familia diversa de péptidos, polipéptidos y proteínas relacionados con anticuerpos humanos (por ejemplo, una familia diversa de anticuerpos), y presentan, presentan y expresan, o comprenden colectivamente al menos una porción de la diversidad de la familia, en el que los vectores o paquetes genéticos comprenden secuencias de ADN abigarradas que codifican una CDR3 de cadena pesada (HC) seleccionada del grupo que consiste en:

(a) una CDR3 de HC que tiene una longitud de alrededor de 3 o alrededor de 4 o alrededor de 5 aminoácidos;

(b) una CDR3 de HC que tiene una longitud de alrededor de 23, alrededor de 24, alrededor de 25, alrededor de 26, alrededor de 27, alrededor de 28, alrededor de 29, alrededor de 30, alrededor de 31, alrededor de 32, alrededor de 33, alrededor de 34 o alrededor de 35 aminoácidos (por ejemplo, una longitud de alrededor de 23 a alrededor de 35 aminoácidos); y

(c) una CDR3 de HC que tiene una longitud de alrededor de 6 a alrededor de 20 aminoácidos (por ejemplo, una longitud de alrededor de 6, alrededor de 7, alrededor de 8, alrededor de 9, alrededor de 10, alrededor de 11, alrededor de 12, alrededor de 13, alrededor de 14, alrededor de 15, alrededor de 16, alrededor de 17, alrededor de 18, alrededor de 19, o alrededor de 20 aminoácidos);

en el que la CDR3 de HC comprende aminoácidos de una región D (por ejemplo, una región D diversificada) (o

fragmento de la misma (por ejemplo, 3 o más aminoácidos de la región D, por ejemplo región D diversificada)) o una región JH (por ejemplo, una región JH extendida).

En algunas realizaciones, la CDR de HC está enriquecida en Tyr (Y) y Ser (S) (por ejemplo, más del 40% de los restos de la CDR3 de HC son Y y/o S).

- 5 En algunas realizaciones, la biblioteca (por ejemplo, los vectores o paquetes genéticos de la misma) comprenden una región D o un fragmento de una región D (por ejemplo, en el que la región D está adyacente a una región JH).

En algunas realizaciones, la biblioteca comprende una región JH, por ejemplo una región JH extendida.

En algunas realizaciones, la CDR3 de HC comprende aminoácidos de una región D o un fragmento de una región D (por ejemplo, en el que la región D está adyacente a una región JH).

- 10 En algunas realizaciones, la región D se selecciona del grupo que consiste en D2-2 (RF 2), D2-8 (RF 2), D2-15 (RF 2), D2-21 (RF 2), D3-16 (RF 2), D3-22 (RF 2), D3-3 (RF-2), D3-9 (RF 2), D3-10 (RF 2), D1-26 (RF 3), D4-11 (RF 2), D4-4 (RF 2), D5-5 (RF 3), D5-12 (RF 3), D5-18 (RF 3), D6-6 (RF1), D6-13 (RF 1), y D6-19 (RF 1).

En algunas realizaciones, la CDR3 de HC comprende aminoácidos de una región JH. La región JH puede ser una región JH extendida. En algunas realizaciones, la región JH extendida se selecciona del grupo que consiste en JH1, JH2, JH3, JH4, JH5, y JH6. En algunas realizaciones, la región JH puede estar enriquecida en restos Y y/o S, por ejemplo puede contener más de alrededor de 40% (por ejemplo, entre alrededor de 43% y alrededor de 80%; por ejemplo, más de alrededor de 40% pero menos de alrededor de 100%) de restos Y y/o S.

- 15 En algunas realizaciones, la región D comprende uno o más restos de cisteína (Cys), y en algunas realizaciones, el uno o más restos de Cys se mantienen constantes (por ejemplo, no varían).
- 20 En algunas realizaciones, la CDR3 de HC (por ejemplo, el ADN que codifica la CDR3 de HC) comprende uno o más codones de relleno entre FR3 y la región D, y cada codón de relleno es individualmente NNK, TMY, TMT, o TMC (TMY, TMT, o TMC codifican S o Y).

En algunas realizaciones, la CDR3 de HC (por ejemplo, el ADN que codifica la CDR3 de HC) comprende uno o más codones de relleno entre la región D y JH, y cada codón de relleno es individualmente NNK, TMY, TMT, o TMC.

- 25 En algunas realizaciones, la biblioteca (por ejemplo, los vectores con paquetes genéticos de la biblioteca) comprende además una CDR1 de HC, CDR2 de HC, y/o una cadena ligera, y también comprende diversidad en la CDR1 de HC, CDR2 de HC, o la cadena ligera comprende diversidad en CDR1 de HC y/o CDR2 de HC, y/o una cadena ligera (por ejemplo, cadena ligera kappa o lambda) (respectivamente). Por ejemplo, la diversidad de CDR3 de HC se puede construir sobre la diversidad de CDR1 de HC, CDR2 de HC, y/o cadenas ligeras. Por ejemplo, la diversidad de la cadena ligera puede estar codificada en la misma molécula de ADN que la diversidad de HC, o las diversidades de LC y HC se pueden codificar en moléculas de ADN distintas.
- 30

En algunos aspectos, la descripción se refiere a una biblioteca que comprende una CDR3 de HC que tiene 3, 4, ó 5 aminoácidos de longitud, en la que la CDR3 comprende aminoácidos de una región JH (por ejemplo, región JH extendida) o de una región D (por ejemplo, una región D diversificada) (o su fragmento (por ejemplo, 3 o más aminoácidos de la región D, por ejemplo región D diversificada)) unida a la porción FR4 de una región JH.

- 35 En algunas realizaciones, la CDR3 de HC es de una región D unida a la porción FR4 de una región JH, y comprende un trímero, un tetrámero, o un pentámero, en el que el trímero, tetrámero, o pentámero no comprende un resto de cisteína.

- 40 En algunas realizaciones, la CDR3 de HC es de una región D unida a la porción FR4 de una región JH, y comprende un trímero, un tetrámero, o un pentámero, en el que el trímero, tetrámero, o pentámero no comprende un codón de parada.

En algunas realizaciones, la región D (por ejemplo, el ADN que codifica la región D) comprende un codón TAG, y el codón TAG es sustituido por un codón seleccionado del grupo que consiste en TCG, TTG, TGG, CAG, AAG, TAT, y GAG.

- 45 En algunas realizaciones, la región D (por ejemplo, el ADN que codifica la región D) comprende un codón TAA, y el codón TAA es sustituido por un codón seleccionado del grupo que consiste en TCA, TTA, CAA, AAA, TAT, y GAA.

En algunas realizaciones, la región D (por ejemplo, el ADN que codifica la región D) comprende un codón TGA, y el codón TGA es sustituido por un codón seleccionado del grupo que consiste en TGG, TCA, TTA, AGA, y GGA.

- 50 En algunas realizaciones, la biblioteca comprende además diversidad en CDR1 de HC y/o CDR2 de HC, y/o una cadena ligera (por ejemplo, cadena ligera kappa o lambda). Por ejemplo, la diversidad de CDR3 de HC se puede construir sobre la diversidad en CDR1 de HC, CDR2 de HC, y/o cadenas ligeras. Por ejemplo, la diversidad de las cadenas ligeras puede estar codificada en la misma molécula de ADN que la diversidad de HC, o las diversidades

de LC y HC pueden estar codificadas en moléculas de ADN distintas.

En algunos aspectos, la descripción proporciona un método para diversificar una biblioteca, comprendiendo el método mutagenizar una biblioteca descrita aquí.

En algunas realizaciones, la mutagenización comprende PCR con tendencia a error.

5 En algunas realizaciones, la mutagenización comprende titubeo.

En algunas realizaciones, la mutagenización comprende titubeo digital.

En algunas realizaciones, la mutagenización introduce, en promedio, alrededor de 1 a alrededor de 10 mutaciones (por ejemplo, alrededor de 1, alrededor de 2, alrededor de 3, alrededor de 4, alrededor de 5, alrededor de 6, alrededor de 7, alrededor de 8, alrededor de 9, alrededor de 10 mutaciones; por ejemplo, cambios de bases) por CDR3 de HC.

10 “Titubeo” es un método para obtener ADN variegado, de manera que se favorece una secuencia original. Si la secuencia original tuvo, por ejemplo, una Ala que podría ser codificada con GCT, la mezcla (0,7 G, 0,1 A, 0,1 T, 0,1 C) se puede usar para la primera posición, (0,7 C, 0,1 A, 0,1 T, 0,1 G) en la segunda posición, y (0,7 T, 0,1 A, 0,1 G, 0,1 C) en la tercera. Se pueden usar otras relaciones de “dopaje”. Esto permite que Ala aparezca alrededor de 50% de las veces, mientras que V, D, G, T, P, y S aparecen alrededor de 7% de las veces. Otros tipos AA se producen a una menor frecuencia.

En algunos aspectos, la presente descripción se extrae, por ejemplo, para mantener un repertorio de CDR1-2 de HC (purificado) y construir diversidad sintética de CDR3 de HC y LC.

20 En algunas realizaciones, la descripción proporciona un casete para presentar una CDR de cadena pesada (HC) titubeada, por ejemplo en el que el casete comprende el casete mostrado en la Tabla 400.

En algunos aspectos, la presente descripción se refiere a una biblioteca en la que los niveles de Tyr están controlados en la CDR3 de HC. En algunas realizaciones, las regiones de CDR3 de HC contienen alrededor de 15% o más (por ejemplo, alrededor de 16%, alrededor de 18%, alrededor de 20%, o más) restos de Tyr. En algunas realizaciones, se insertan niveles elevados (por ejemplo, más de alrededor de 20%) de Tyr en la CDR3 de HC de los miembros de la biblioteca, por ejemplo en las regiones D y muñones J (o secuencias sintéticas correspondientes a ellos) que contienen Tyr. En algunas realizaciones, en posiciones directrices o de relleno DJ (o secuencias sintéticas que corresponden a ellas), se permite Tyr, pero no más de 20%. En algunas realizaciones, las regiones de CDR3 de HC contienen menos de alrededor de 15% (por ejemplo, alrededor de 14%, alrededor de 12%, alrededor de 10%, alrededor de 8%, alrededor de 6% o menos) de restos de Tyr. En algunas realizaciones, las posiciones directrices de HC o de relleno de DJ (o secuencias sintéticas correspondientes a ellas) contienen menos de alrededor de 15% (por ejemplo, alrededor de 14%, alrededor de 12%, alrededor de 10%, alrededor de 8%, alrededor de 6% o menos) de restos de Tyr.

35 En algunos aspectos, la descripción se refiere a una biblioteca de paquetes genéticos que codifica una cadena pesada de anticuerpo humano en la que una secuencia de aminoácidos progenitora comprende una secuencia de VH seguida de cero a diez aminoácidos seleccionados del grupo que consiste en (Y, S, D, L, R), seguido de una región D humana o fragmento de una región D, seguido de cero a diez aminoácidos seleccionados del grupo que consiste en (Y, S, R, D, L), seguido de un segmento JH que comprende al menos W103 en adelante, en el que el ADN variable que codifica esta secuencia se sintetiza de manera que la secuencia de aminoácidos progenitora es la más probable (por ejemplo, mediante titubeo).

40 En algunos aspectos, la descripción se refiere a una biblioteca de cadenas ligeras que tienen regiones de marco de línea germinal y en la que las CDR se varían de manera que los restos remotos del sitio de combinación o que tienen grupos laterales escondidos se mantienen constantes. En algunas realizaciones, se usa un método de síntesis de ADN variable de manera que la secuencia de línea germinal es la más probable (por ejemplo, mediante titubeo).

45 En algunos aspectos, la descripción se refiere a una biblioteca de miembros diversos que codifican regiones variables de unión a antígeno como se describe aquí.

En algunos aspectos, la descripción se refiere a una biblioteca de miembros diversos que codifican regiones de CDR3 de HC como se describe aquí. En algunas realizaciones, la biblioteca es una biblioteca de la Tabla 1097.

50 En algunos aspectos, la descripción se refiere a una biblioteca de miembros diversos, codificando cada miembro una CDR3 de HC, en el que

al menos 1, 2, 3, 4, 5, 6, 7, u 8 posiciones en la HDCR3, respectivamente, están ocupadas por G, S, R, D, L, e Y en la biblioteca en las siguientes proporciones [1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y] y opcionalmente,

las últimas 4 posiciones de CDR3 de HC están representadas según lo siguiente:

el aminoácido progenitor está presente 3, 4, 5, 6, 7, 8, 10 veces con igual probabilidad que otros tipos de aminoácidos, en el que los otros tipos de aminoácidos comprenden Y, S, D, R, G.

En algunos aspectos, la descripción se refiere a una biblioteca de miembros diversos, codificando cada miembro una CDR3 de HC, en el que

- 5 al menos una y preferiblemente todas las primeras 1, 2, 3, 4, 5, 6, 7, u 8 posiciones en la CDR3 de HC están ocupadas por G, S, R, D, L, e Y en la biblioteca en las siguientes proporciones [1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y] y opcionalmente,

las últimas 4 posiciones de HCDR3 están representadas según lo siguiente:

- 10 el aminoácido progenitor está presente 3, 4, 5, 6, 7, 8, 10 veces con igual probabilidad que otros tipos de aminoácidos, en el que los otros tipos de aminoácidos comprenden Y, S, D, R, G.

En algunos aspectos, la descripción se refiere a una biblioteca de miembros diversos, codificando cada miembro una CDR3 de HC, en el que,

la longitud de CDR3 de HC es 10, 11, ó 12 posiciones;

- 15 cada una de las primeras 6, 7 u 8 posiciones en la CDR3 de HC, respectivamente, está ocupada por G, S, R, D, L, e Y en la biblioteca en las siguientes proporciones [1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y];

las últimas 4 posiciones de HCDR3 están representadas según lo siguiente:

el aminoácido progenitor está presente 3, 4, 5, 6, 7, 8, 10 veces con igual probabilidad que otros tipos de aminoácidos, en el que los otros tipos de aminoácidos comprenden Y, S, D, R, G.

- 20 En algunas realizaciones, cada una de las últimas 4 posiciones de CDR3 de HC está representada en la biblioteca como 7/12 progenitor, más 1/12 cada uno de Y, S, D, R, y G.

En algunas realizaciones, cada una de las últimas 4 posiciones de CDR3 de HC está representada en la biblioteca como A6= 7/12 A, más 1/12 cada uno de Y, S, D, R, y G; F7 = 7/12 F más 1/12 cada uno de Y, S, D, R, y G; D8 = 7/11 D más 1/11 de Y, S, R, y G; I9 = 7/12 I más 1/12 Y, S, R, D, G.

En algunas realizaciones, los miembros codifican además CDR1 de HC, CDR2 de HC.

- 25 En algunas realizaciones, los miembros codifican además regiones 1-4 del marco (FR).

En algunas realizaciones, los miembros codifican CDR1 de HC, CDR2 de HC y regiones 1-4 de FR.

En algunas realizaciones, los miembros comprenden un marco de 3-23 HC.

En algunas realizaciones, la biblioteca comprende además una región variable de LC.

En algunas realizaciones, la biblioteca comprende miembros que codifican regiones variables de LC diversas.

- 30 En algunas realizaciones, los miembros que comprenden una región variable de LC comprenden un marco de LC A27.

En algunas realizaciones, la biblioteca es una biblioteca de presentación, por ejemplo una biblioteca de presentación de fagos.

- 35 En algunas realizaciones, la biblioteca comprende tiene al menos 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} miembros diversos.

En algunos aspectos, la descripción se refiere a un método para seleccionar un miembro de la biblioteca, que comprende poner en contacto una biblioteca descrita aquí con una diana, permitir que un miembro se una a dicha diana, y recuperar el miembro que se une a la diana.

- 40 Estas realizaciones de la presente invención, otras realizaciones, y sus rasgos y características serán manifiestos a partir de la descripción, dibujos, y reivindicaciones que siguen.

DESCRIPCIÓN DETALLADA

- 45 Los anticuerpos ("Ab") concentran su diversidad en aquellas regiones que están implicadas en la determinación de la afinidad y especificidad del Ab para dianas específicas. Estas regiones pueden ser diversas en secuencia o en longitud. Generalmente, son diversas en ambos aspectos. Sin embargo, dentro de las familias de anticuerpos humanos, las diversidades, tanto en secuencia como en longitud, no son verdaderamente aleatorias. Más bien,

algunos restos de aminoácidos se prefieren en determinadas posiciones de las CDRs, y se prefieren algunas longitudes de las CDRs. Estas diversidades preferidas explican la diversidad natural de la familia de anticuerpos.

Según esta invención, y como se describe de forma más completa más abajo, se pueden preparar y usar bibliotecas de vectores y paquetes genéticos que codifican miembros de una familia diversa de anticuerpos humanos que comprenden CDR3s de cadena pesada (HC) que tienen una longitud entre alrededor de 3 y alrededor de 35 aminoácidos. En ciertas realizaciones, las CDR3s de HC también pueden ser ricas en Y y S, y/o comprender regiones D diversificadas. También se proporcionan bibliotecas focalizadas que comprenden tales CDR3s de HC.

Cuando una célula inmunitaria construye una cadena pesada de anticuerpo, conecta un segmento V a un segmento D, y éste a un segmento J. El segmento D es opcional, y alrededor del 50% de los Abs humanos tienen Ds reconocibles. La célula puede llevar a cabo una edición considerable en los sitios de unión (V a D, D a J, o V a J) tanto eliminando como añadiendo bases, pero no exactamente al azar. El anticuerpo inicialmente reorganizado se presenta en la superficie de la célula y, si se une a un antígeno (Ag), se estimula a la célula a llevar a cabo mutaciones somáticas para mejorar la afinidad. Hay puntos calientes codificados en los genes de línea germinal de inmunoglobulinas, de manera que es muy probable que ciertos lugares en el gen de Ab pase a través de un conjunto particular de mutaciones en busca de un mejor ligador a un Ag persistente. En la naturaleza, algunas de las mutaciones están en posiciones del marco, pero la mayoría están en las regiones determinantes de la complementariedad (CDRs). Es de particular interés la CDR3 de la cadena pesada (HC), debido a que muestra no sólo un grado elevado de diversidad de secuencia sino también diversidad de longitud. Se han construido bibliotecas de anticuerpos (Ab) en las que las CDRs están sustituidas por ADN aleatorio, y se han obtenido Abs útiles. Sin embargo, algunos Abs terapéuticos muestran un grado significativo de antigenicidad. Es posible que los Abs que estén más próximos a la línea germinal humana fuesen menos antigénicos.

Definiciones

Por conveniencia, antes de describir adicionalmente la presente invención, se definen aquí ciertos términos empleados en la memoria descriptiva, en los ejemplos y en las reivindicaciones anejas.

Las formas singulares “un”, “una”, y “el/la” incluyen referencias plurales excepto que el contexto dicte claramente otra cosa.

El término “afinidad” o “afinidad de unión” se refiere a la constante de asociación aparente o K_a . La K_a es el recíproco de la constante de disociación (K_d). Una proteína de unión puede tener, por ejemplo, una afinidad de unión de al menos 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} y 10^{11} M^{-1} para una molécula diana particular. Una mayor afinidad de unión de una proteína de unión a una primera diana con respecto a una segunda diana se puede indicar mediante una mayor K_a (o un valor numérico más pequeño de K_D) para la unión de la primera diana que la K_a (o valor numérico K_D) para la unión a la segunda diana. En tales casos, la proteína de unión tiene especificidad por la primera diana (por ejemplo, una proteína en una primera conformación o imitación de la misma) con respecto a la segunda diana (por ejemplo, la misma proteína en una segunda conformación o imitación de la misma, o una segunda proteína). Las diferencias en la afinidad de unión (por ejemplo, para especificidad u otras comparaciones) puede ser de al menos 1,5, 2, 3, 4, 5, 10, 15, 20, 37,5, 50, 70, 80, 91, 100, 500, 1000, o 10^5 veces.

La afinidad de unión puede determinarse mediante una variedad de métodos que incluyen diálisis en el equilibrio, unión en el equilibrio, filtración en gel, ELISA, resonancia de plasmones superficiales, o espectroscopia (por ejemplo, usando un ensayo de fluorescencia). Las condiciones ejemplares para evaluar la afinidad de unión son el amortiguador TRIS (TRIS 50 mM, NaCl 150 mM, $CaCl_2$ 5 mM a pH 7,5). Estas técnicas pueden usarse para medir la concentración de proteína de unión enlazada y libre en función de la concentración de la proteína de unión (o diana). La concentración de la proteína de unión enlazada ([Enlazada]) está relacionada con la concentración de la proteína de unión libre ([Libre]) y la concentración de sitios de unión para la proteína de unión en la diana, en la que (N) es el número de sitios de unión por molécula diana, mediante la siguiente ecuación:

$$[Enlazada] = N \cdot [Libre] / ((1/K_A) + [Libre]).$$

No siempre es necesario hacer una determinación exacta de K_A , aunque, ya que algunas veces es suficiente obtener una medida cuantitativa de la afinidad, por ejemplo determinada usando un método tal como ELISA o análisis FACS, es proporcional a K_A , y de este modo se puede utilizar para comparaciones, tales como la determinación de si una mayor afinidad es, por ejemplo, 2 veces mayor, para obtener una medida cualitativa de la afinidad, o para obtener una inferencia de la afinidad, por ejemplo mediante la actividad en un ensayo funcional, por ejemplo un ensayo *in vitro* o *in vivo*.

El término “anticuerpo” se refiere a una proteína que incluye al menos un dominio variable de inmunoglobulina o una secuencia del dominio variable de inmunoglobulina. Por ejemplo, un anticuerpo puede incluir una región variable de cadena pesada (H) (abreviada en este documento como VH), y una región variable de cadena ligera (L) (abreviada en este documento como VL). En otro ejemplo, un anticuerpo incluye dos regiones variables de cadena pesada (H) y dos regiones variables de cadena ligera (L). La cadena pesada y la cadena ligera también se pueden abreviar como HC y LC, respectivamente. El término “anticuerpo” abarca fragmentos de unión a antígeno de los anticuerpos (por ejemplo, anticuerpos monocaténarios, fragmentos Fab y sFab, $F(ab')_2$, fragmentos Fd, fragmentos Fv, scFv, y

fragmentos de anticuerpos de dominio (dAb) (de Wildt et al., Eur J Immunol. 1996; 26 (3): 629-39)), así como anticuerpos completos. Un anticuerpo puede tener las características estructurales de IgA, IgG, IgE, IgD, IgM (así como sus subtipos). Los anticuerpos pueden proceder de cualquier fuente, pero se prefieren de primate (primate humano y no humano) y primatizados.

- 5 Las regiones VH y VL pueden subdividirse adicionalmente en regiones de hipervariabilidad, denominadas "regiones determinantes de la complementariedad" ("CDR"), intercaladas con regiones que están más conservadas, denominadas "regiones del marco" ("FR"). La extensión de la región marco y las CDRs se ha definido con precisión (véase, Kabat, EA, et al. (1991) Sequences of Proteins of Immunological Interest, Quinta Edición, Departamento de Salud y Servicios Humanos de los Estados Unidos, Publicación del NIH No. 91-3242, y Chothia, C. et al. (1987) J. Mol. Biol. 196: 901-917, véase también www.hgmp.mrc.ac.uk). Se usan aquí las definiciones de Kabat. Cada VH y VL se compone típicamente de tres CDRs y cuatro FRs, dispuestas desde el término amino hasta el término carboxi en el siguiente orden: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4.

- 15 La cadena VH o VL del anticuerpo puede incluir además todo o parte de una región constante de cadena pesada o ligera, para formar de ese modo una cadena de inmunoglobulina pesada o ligera, respectivamente. En una realización, el anticuerpo es un tetrámero de dos cadenas pesadas de inmunoglobulina y dos cadenas ligeras de inmunoglobulina, en el que las cadenas pesada y ligera de inmunoglobulina están interconectadas, por ejemplo, mediante enlaces de disulfuro. En las IgGs, la región constante de cadena pesada incluye tres dominios de inmunoglobulina, CH1, CH2 y CH3. La región constante de cadena ligera incluye un dominio CL. La región variable de las cadenas pesada y ligera contiene un dominio de unión que interactúa con un antígeno. Las regiones constantes de los anticuerpos median típicamente la unión del anticuerpo a tejidos o factores hospedantes, incluyendo diversas células del sistema inmunitario (por ejemplo, células efectoras) y el primer componente (C1q) del sistema clásico del complemento. Las cadenas ligeras de la inmunoglobulina pueden ser de los tipos kappa o lambda. En una realización, el anticuerpo está glicosilado. Un anticuerpo puede ser funcional para citotoxicidad dependiente del anticuerpo y/o citotoxicidad mediada por el complemento.

- 25 Una o más regiones de un anticuerpo pueden ser humanas o efectivamente humanas. Por ejemplo, una o más de las regiones variables pueden ser humanas o efectivamente humanas. Por ejemplo, una o más de las CDRs pueden ser humanas, por ejemplo CDR1 de HC, CDR2 de HC, CDR3 de HC, CDR1 de LC, CDR2 de LC, y CDR3 de LC. Cada una de las CDRs de cadena ligera puede ser humana. CDR3 de HC puede ser humana. Una o más de las regiones del marco pueden ser humanas, por ejemplo FR1, FR2, FR3, y FR4 de la HC o LC. Por ejemplo, la región de Fc puede ser humana. En una realización, todas las regiones del marco son humanas, por ejemplo derivadas de una célula somática humana, por ejemplo una célula hematopoyética que produce inmunoglobulinas o una célula no hematopoyética. En una realización, las secuencias humanas son secuencias de línea germinal, por ejemplo codificadas por un ácido nucleico de línea germinal. En una realización, los restos del marco (FR) de un Fab seleccionado se pueden convertir al del tipo de aminoácido del resto correspondiente en el gen de línea germinal de primate más similar, especialmente el gen de línea germinal humana. Una o más de las regiones constantes pueden ser humanas o efectivamente humanas. Por ejemplo, al menos 70, 75, 80, 85, 90, 92, 95, 98, o 100% de un dominio variable de inmunoglobulina, la región constante, los dominios constantes (CH1, CH2, CH3, CL), o el anticuerpo completo pueden ser humanos o efectivamente humanos.

- 40 Todo o parte de un anticuerpo puede estar codificado por un gen de inmunoglobulina o un segmento del mismo. Ejemplos de genes de inmunoglobulina humana incluyen los genes de región constante kappa, lambda, alfa (IgA1 e IgA2), gamma (IgG1, IgG2, IgG3, IgG4), delta, épsilon y mu, así como los muchos genes de región variable de inmunoglobulina. Las "cadenas ligeras" de longitud completa de inmunoglobulina (alrededor de 25 kDa o alrededor de 214 aminoácidos) son codificadas por un gen de la región variable en el término NH2 (alrededor de 110 aminoácidos) y un gen de la región constante kappa o lambda en el término COOH. Las "cadenas pesadas" de longitud completa de inmunoglobulina (alrededor de 50 kDa o alrededor de 446 aminoácidos) están codificadas de manera similar por un gen de región variable (alrededor de 116 aminoácidos) y uno de los otros genes de región constante mencionados anteriormente, por ejemplo gamma (que codifica alrededor de 330 aminoácidos). La longitud de HC humana varía considerablemente debido a que CDR3 de HC varía desde alrededor de 3 restos de aminoácidos a más de 35 restos de aminoácidos.

- 50 Aquí, las expresiones "segmento D" y "región D" se usan de forma intercambiable y son idénticas. Se ha de entender que estas definiciones tienen tanto representaciones de ADN como de aminoácidos, y que lo que se quiere decir está claro a partir del contexto.

- 55 Una "biblioteca" o "biblioteca de presentación" se refiere a una colección de secuencias nucleotídicas, por ejemplo ADN, en los clones; o una colección genéticamente diversa de polipéptidos presentados en paquetes de presentación replicables capaces de selección o identificación para proporcionar un polipéptido individual o una población mixta de polipéptidos.

El término "paquete", como se usa aquí, se refiere a un paquete de presentación genético replicable en el que la partícula presenta un polipéptido en su superficie. El paquete puede ser un bacteriófago que presenta un dominio de unión a antígeno en su superficie. Este tipo de paquetes se ha denominado un anticuerpo fágico (pAb).

Una "diana predeterminada" se refiere a una molécula diana cuya identidad es conocida antes de usarla en cualquiera de los métodos descritos.

La expresión "paquete de presentación replicable", como se usa aquí, se refiere a una partícula biológica que tiene información genética que proporciona a la partícula con la capacidad para replicarse. La partícula puede presentar sobre su superficie al menos parte de un polipéptido. El polipéptido puede estar codificado mediante información genética nativa a la partícula y/o colocada artificialmente en la partícula o un ancestro de ella. El polipéptido presentado puede ser cualquier miembro de un par de unión específica, por ejemplo dominios de cadenas pesadas o ligeras basados en una molécula inmunoglobulínica, una enzima o un receptor, etc. La partícula puede ser, por ejemplo, un virus, por ejemplo un bacteriófago tal como fd o M13.

El término "vector" se refiere a una molécula de ADN capaz de replicarse en un organismo hospedante, en el que se inserta un gen para construir una molécula de ADN recombinante. Un "vector fágico" es un vector derivado mediante modificación de un genoma de fago, que contiene un origen de replicación para un bacteriófago, pero ninguno para un plásmido. Un "vector fagómido" es un vector derivado mediante modificación un genoma plasmídico, que contiene un origen de replicación para un bacteriófago así como el origen de replicación plasmídico.

Al discutir los oligonucleótidos, la notación "[RC]" indica que el complemento inverso del oligonucleótido mostrado es el que se va a usar.

CRD3s de cadena pesada de anticuerpo humano

El gen de línea germinal (GLG) de cadena pesada ("HC") 3-23 (también conocido como VP-47) da cuenta de alrededor de 12% de todos los Abs humanos, y se prefiere como el marco en la realización preferida de la invención. Sin embargo, se debería entender que también se pueden usar otros marcos bien conocidos, tales como 4-34, 3-30, 3-30.3 y 4-30.1, sin separarse de los principios de las diversidades focalizadas de esta invención.

Además, JH4 (YFDYW₁₀₃GQGTLTVSS (SEQ ID NO:1)) aparece más a menudo que JH3 en anticuerpos nativos. Por tanto, se prefiere para las bibliotecas focalizadas de esta invención. Sin embargo, igualmente se podrían usar JH3 (AFDIW₁₀₃GQGTMTVSS (SEQ ID NO:2)), JH6 (YYYYYGMDW₁₀₃GQGTTVTVSS (SEQ ID NO:3)), JH1, JH2, o JH5. JH2 tiene la ventaja de tener RG en 105-106 en lugar de QG en todos los otros JHs humanos. JH3 tiene la desventaja de M₁₀₈. En una colección de 1419 Abs que fueron positivos para ELISA para al menos una diana, se observaron 17 JH1s, 31 JH2s, 452 JH3s, 636 JH4s, 32 JH5s, y 251 JH6s. Si están presentes, las porciones subrayadas dobles de los JHs se consideran parte de CDR3. En la Tabla 3, las partes de FR4 de los JHs están subrayadas.

La frecuencia a la que cada aminoácido apareció en las CDR3s de HC de estos 1419 Abs se tabuló y se registró en la Tabla 75. Obsérvese que el aminoácido más común es Tyr, siguiéndole Gly, Asp, Ser, y Arg en ese orden. Rel. Up es la abundancia relativa de cada tipo en comparación con Cys, el más común. Rel. Down es la abundancia de cada tipo en comparación con Tyr, el más común. Por tanto, los tipos de aminoácidos preferidos para la sustitución en las CDR3s de HC son Y, G, D, S, y R.

Naturalmente, las CDR3s de HC varían en longitud. Alrededor de la mitad de las HCs humanas consisten en los componentes: V::nz::D::ny::JHn en el que V es un gen V, nz es una serie de bases que son esencialmente aleatorias, D es un segmento D, a menudo con edición importante en ambos extremos, ny es una serie de bases que son esencialmente aleatorias, y JHn es uno de los seis segmentos de JH, a menudo con edición importante en el extremo 5'. Los segmentos D parecen proporcionar segmentos espaciadores que permiten el plegamiento de la IgG. La mayor diversidad está en las uniones de V con D y de D con JH.

Corbett et al. (Corbett SJ, Tomlinson IM, Sonnhhammer EL, Buck D, Winter G. J Mol Biol. 1997 V270:587-97) mostraron que el sistema inmunitario humano no inserta múltiples segmentos D y segmentos D recombinantes. No obstante, los segmentos D se han seleccionado para ser buenos componentes de las CDR3s de HC, y la presente invención comprende CDR3 de HC que contiene más de un segmento D.

Los segmentos D humanos tienen algunas desviaciones muy fuertes. El cómputo de los 522 aminoácidos en los segmentos D humanos es Y 70 (13,4%), L 63 (12,1%), V 52 (10%), G 49 (9,4%), I 41 (7,9%), T 40 (7,7%), S 33 (6,3%), W 27 (5,2%), D 21 (4%), A 19 (3,6%), R 16 (3,1%), TAG 15 (2,9%), N 14 (2,7%), Q 11 (2,1 %), C 9 (1,7%), E 9 (1,7%), F 8 (1,5%), M 8 (1,5%), TGA 8 (1,5%), TAA 7 (1,3%), P 1 (0,2%), H 1 (0,2%), y K 0 (0%). Hay un D (2-8 RF 1) que tiene una Cys sin emparejar pero también un codón de parada TGA, de manera que se usó poco. De este modo, los segmentos D son principalmente hidrófobos. Las frecuencias de aminoácidos en las CDR3s de HC humanas se muestran en la Tabla 75. Hay tanto similitudes como diferencias en las frecuencias. En el conjunto de las CDR3s de HC, Tyr es el más común y sólo se aproxima Gly (96% tan común como Tyr). Asp (75% tan común como Tyr), Ser (53% tan común como Tyr). Leu, Val, e Ile son relativamente comunes en los segmentos D si todos los segmentos D se cuentan como iguales. El sistema inmunitario no usa los segmentos D con igual frecuencia. La Tabla 77 muestra la frecuencia de utilización de los segmentos D. Los segmentos D que se usan más a menudo son muy ricos en Tyr, Gly, Ser, y Asp. Arg no se encuentra en los segmentos D más usados a menudo, ni Arg es codificado en ninguna de las porciones de CDR de los segmentos de JH. Arg se hace prominente mediante mutación de V, D, y J, o en las regiones de relleno entre V y D, D y J, o V y J. En esta muestra, el 50% de todos los

aminoácidos son Tyr, Gly, Asp, Ser, o Arg. En una realización de la presente invención, las sustituciones de las secuencias de CDR3 de HC "progenitoras" están limitadas al conjunto de aminoácidos que consiste en Tyr, Gly, Ser, Asp, y Arg. En una realización de la presente invención, Arg se hace común en las regiones de relleno entre V y D, entre D y J, o entre V y J.

- 5 En las bibliotecas preferidas de esta invención, se usan ambos tipos de CDR3s de HC. En las CDR3s de HC que no tienen un segmento D identificable, la estructura es V::nz::JHn (n = 1,6), en la que JH se edita habitualmente en el extremo 5. En las CDR3s de HC que tienen un segmento D identificable, la estructura es V::nz::D::ny::JHn.

- Se proporcionan aquí CDR3s de HC que tienen una longitud entre alrededor de 3 y alrededor de 35 aminoácidos. Las CDR3s de HC también pueden ser ricas, en ciertas realizaciones, en Y y S y/o comprender regiones D diversificadas, cuando está presente una región D. Por ejemplo, las CDR3s de HC pueden contener entre alrededor de 43% y alrededor de 80% de restos de Y y/o S, por ejemplo, alrededor de 43%, alrededor de 48%, alrededor de 69%, alrededor de 63%, alrededor de 71%, alrededor de 62%, alrededor de 58%, alrededor de 68%, alrededor de 80%, alrededor de 77%, o más de alrededor de 40%, o alrededor de 40% a menos de alrededor de 100%, de los restos son Y y/o S. Por ejemplo, no todos los restos en la CDR3 son Y y/o S. Las CDR3s de HC pueden comprender, en ciertas realizaciones, una región JH extendida. Los diseños de componentes de CDR3 de HC ejemplares de las bibliotecas preferidas de esta invención se muestran y describen en los Ejemplos 1, 2 y 3.

- En algunas realizaciones, la diversidad (por ejemplo, en una CDR, por ejemplo CDR3 de HC, o región del marco (por ejemplo, región del marco cerca o adyacente a CDR, por ejemplo, CDR3, por ejemplo, CDR3 de HC) se genera para crear, en promedio, alrededor de 1, alrededor de 2, alrededor de 3, alrededor de 4, alrededor de 5, alrededor de 6, alrededor de 7, alrededor de 8, alrededor de 9, alrededor de 10, o alrededor de 1 a alrededor de 10 mutaciones (por ejemplo, cambio de bases), por ejemplo, por CDR (por ejemplo, CDR3 de HC) o región del marco (por ejemplo, región del marco cerca o adyacente a CDR, por ejemplo, CDR3, por ejemplo, CDR3 de HC). En algunas implementaciones, la mutagénesis está dirigida a regiones conocidas o que probablemente estén en la interfaz de unión. Además, la mutagénesis se puede dirigir hacia regiones del marco cerca o adyacentes a las CDRs. En el caso de anticuerpos, la mutagénesis también puede limitarse a una o a unas pocas de las CDRs, por ejemplo, para realizar mejoras precisas por etapas. Igualmente, si los ligandos identificados son enzimas, la mutagénesis puede proporcionar anticuerpos que son capaces de unirse al sitio activo y a la vecindad. La región de CDR o del marco (por ejemplo, una CDR3 de HC descrita aquí) puede someterse, en ciertas realizaciones, a PCR con tendencia a error, para generar la diversidad. Este enfoque usa una versión "descuidada" de PCR, en la que la polimerasa tiene una tasa de error bastante alto (hasta 2%), para amplificar la secuencia de tipo salvaje, y se describe generalmente en Pritchard, et al. (2005) J. Theor. Biol. 234: 497-509 y Leung et al. (1989) Technique 1:11-15. Otras técnicas de mutagénesis ejemplares incluyen el barajado del ADN usando escisión aleatoria (Stemmer (1994) Nature 389-391; denominado "barajado del ácido nucleico"), RACHITT™ (Coco et al. (2001) Nature Biotech. 19:354), mutagénesis dirigida al sitio (Zoller et al. (1987) Nucl Acids Res 10:6487-6504), mutagénesis de casetes (Reidhaar-Olson (1991) Methods Enzymol. 208:564-586) e incorporación de oligonucleótidos degenerados (Griffiths et al. (1994) EMBO J. 13:3245).

En algunas realizaciones de la invención, se escogen los segmentos D en los que una mayoría de los restos son Ser o Tyr. En algunas realizaciones, cuando se sintetiza el ADN que codifica la región D, cada resto de Ser o Tyr es codificado por TMT, TMC, o TMY de manera que el aminoácido codificado es Ser o Tyr.

- 40 En algunas realizaciones, las secuencias de CDR3 de HC descritas aquí se pueden someter a selección para marcos de lectura abiertos fusionando la secuencia que codifica la CDR3 de HC de interés en el marco a un gen de resistencia a antibiótico, tal como el gen *Kan^R*, y seleccionando para resistencia a canamicina. Las células en las que la CDR3 potencial tiene un codón de parada o un desplazamiento del marco no tendrán la resistencia a antibiótico, y esa secuencia se eliminará.

- 45 Métodos de construcción de bibliotecas que comprenden CDR3s de cadena pesada de anticuerpo humano y bibliotecas que comprenden CDR3s de cadena pesada de anticuerpo humano

- Una biblioteca de anticuerpos es una colección de proteínas que incluye proteínas que tienen al menos una secuencia de dominio variable de inmunoglobulina. Por ejemplo, se pueden usar dominios variables camelizados (por ejemplo, dominios VH) como un andamiaje para una biblioteca de proteínas que incluya sólo una secuencia de dominio variable de inmunoglobulina. En otro ejemplo, las proteínas incluyen dos secuencias de dominios variables, por ejemplo un dominio VH y VL, que son capaces de emparejarse. Una biblioteca de anticuerpos se puede preparar a partir de una biblioteca de ácidos nucleicos (una biblioteca que codifica anticuerpos) que incluye secuencias que codifican anticuerpos, por ejemplo que comprenden las secuencias que codifican las CDR3s de HC proporcionadas aquí.

- 55 En los casos en los que se use una biblioteca de presentación, cada miembro de la biblioteca que codifica anticuerpos se puede asociar con el anticuerpo que codifica. En el caso de biblioteca de fagos, la proteína del anticuerpo se asocia físicamente (directa o indirectamente) con una proteína de cubierta de fago. Un miembro de la biblioteca de presentación de anticuerpos típico presenta un polipéptido que incluye un dominio VH y un dominio VL. El miembro de la biblioteca de presentación puede presentar el anticuerpo como un fragmento Fab (por ejemplo,

usando dos cadenas polipeptídicas) o un Fv monocatenario (por ejemplo, usando una sola cadena polipeptídica). También se pueden usar otros formatos.

Como en el caso del Fab y otros formatos, el anticuerpo presentado puede incluir una o más regiones constantes como parte de una cadena ligera y/o pesada. En una realización, cada cadena incluye una región constante, por ejemplo como en el caso de un Fab. En otras realizaciones, se incluyen regiones constantes adicionales. También es posible añadir una o más regiones constantes a una molécula después de que se identifica que tiene un sitio de unión a antígeno útil. Véase, por ejemplo, el documento US 2003-0224408.

Las bibliotecas de anticuerpos se pueden construir mediante un número de procedimientos (véanse, por ejemplo, de Haard et al. (1999) *J. Biol. Chem* 274:18218-30; Hoogenboom et al. (1998) *Immunotechnology* 4:1-20, Hoogenboom et al. (2000) *Immunol Today* 21:371-8, y Hoet et al. (2005) *Nat Biotechnol.* 23(3):344-8.

En ciertas realizaciones para construir bibliotecas, las cadenas pesadas que comprenden las CDR3s descritas aquí y las cadenas ligeras kappa y lambda se construyen mejor en vectores distintos. En primer lugar, se diseña un gen sintético para encarnar cada uno de los dominios variables sintéticos. Las cadenas ligeras se pueden unir mediante sitios de restricción para ApaLI (situado en el mismísimo extremo de la secuencia señal) y AclI (situado después del codón de parada). La cadena pesada se puede unir mediante Sfil (situado en la secuencia señal de Pe1B) y NotI (situado en el ligador entre CH1 y la proteína de anclaje). También se pueden usar secuencias señal distintas de Pe1B, por ejemplo una secuencia señal de M13 pIII.

Los genes iniciales se pueden obtener con secuencias "de relleno" en lugar de las CDRs deseadas. Un "relleno" es una secuencia que se va a cortar y sustituir por ADN diverso, pero que no permite la expresión de un gen de anticuerpo funcional. Por ejemplo, el relleno puede contener varios codones de parada y sitios de restricción que no se producirán en el vector de la biblioteca acabada correcta. Los rellenos se usan para evitar que una secuencia cualquiera de CDR esté muy representada.

La cadena pesada y las cadenas ligeras kappa o lambda se pueden construir en un único vector o en paquetes genéticos (por ejemplo, para presentación o presentación y expresión) que tienen sitios de restricción apropiados que permiten la clonación de estas cadenas. Los procedimientos para construir tales vectores son bien conocidos y se usan ampliamente en la técnica. Preferiblemente, una biblioteca de una cadena pesada y una cadena ligera kappa y una biblioteca de una cadena pesada y una cadena ligera lambda se prepararían por separado.

Muy preferiblemente, la presentación es sobre la superficie de un derivado del fago M13. El vector más preferido contiene todos los genes de M13, un gen de resistencia a antibiótico, y el casete de presentación. El vector preferido se proporciona con sitios de restricción que permiten la introducción y corte de miembros de la familia diversa de genes, como casetes. El vector preferido es estable frente a la reorganización bajo las condiciones de crecimiento usadas para amplificar el fago.

La diversidad capturada por la presente invención se puede presentar y/o expresar en un vector fagómico (por ejemplo, pMID21 (secuencia de ADN mostrada en la Tabla 35)) que presenta y/o expresa el péptido, polipéptido o proteína. Tales vectores se pueden usar también para almacenar la diversidad para la presentación y/o expresión subsiguientes usando otros vectores o fago.

En todavía otras realizaciones, un método denominado la optimización rápida de cadenas ligeras o "ROLIC" descrito en los documentos U.S.S.N 61/028.265 presentado el 13 de 2008, U.S.S.N. 61/043.938 presentado el 10 de abril de 2008, y U.S.S.N. 12/371.000 presentado el 13 de febrero de 2009, se coloca una gran población de LCs en un vector de fago que hace que se presenten en el fago. Se clona una pequeña población (por ejemplo, 3, 10, ó 25) de HCs en *E. coli* de manera que las HCs se segregan en el periplasma, por ejemplo aquellas HCs que tienen las CDR3s descritas aquí. La *E. coli* se infecta entonces con los vectores fágicos que codifican la población grande de LCs para producir los emparejamientos proteicos HC/LC en el fago. Las partículas del fago poseen sólo un gen de LC.

En otro aspecto, en un método denominado la selección económica de cadenas pesadas o "ESCH", también descrito en los documentos U.S.S.N 61/028.265 presentado el 13 de 2008, U.S.S.N. 61/043.938 presentado el 10 de abril de 2008, y U.S.S.N. 12/371.000 presentado el 13 de febrero de 2009, se puede colocar una pequeña población de LCs en un vector que hace que se segreguen. Se construye una nueva biblioteca de HCs en el fago, tal como aquellas proporcionadas aquí que comprenden las CDR3s. Las LCs y HCs se pueden combinar entonces mediante el método de infección mucho más eficiente. Una vez que se selecciona un pequeño conjunto de HC eficaces, éste se puede usar como tal, se puede alimentar a ROLIC para obtener un emparejamiento óptimo de HC/LC, o se puede clonar en una biblioteca de Fab de LCs para la selección clásica.

La diversidad capturada por la presente invención se puede presentar y/o expresar usando un vector adecuado para la expresión en una célula eucariota, por ejemplo un vector de levadura, por ejemplo para la expresión en una célula de levadura.

Otros tipos de presentación de proteínas incluyen la presentación a base de células (véase, por ejemplo, el documento WO 03/029,456); la presentación del ribosoma (véanse, por ejemplo, Mattheakis et al. (1994) *Proc. Natl.*

Acad. Sci. USA 91:9022 y Hanes et al. (2000) Nat Biotechnol. 18:1287-92); fusiones de proteínas con ácidos nucleicos (véase, por ejemplo, la patente U.S. nº 6.207.446); y la inmovilización a una etiqueta no biológica (véase, por ejemplo, la patente U.S. nº 5.874.214).

Los anticuerpos aislados de las bibliotecas de la presente descripción se pueden analizar para determinar el tipo de la LC y el gen de línea germinal más cercano. En una realización preferida, los restos del marco no de línea germinal se cambian nuevamente al aminoácido de línea germinal en tanto que la afinidad de unión y la especificidad no se vean afectadas de forma adversa en un grado inaceptable. Las sustituciones se pueden realizar como un grupo o de forma individual. Las secuencias de línea germinal humana se describen en Tomlinson, I.A. et al., 1992, J. Mol. Biol. 227:776-798; Cook, G. P. et al., 1995, Immunol. Today 16 (5): 237-242; Chothia, D. et al., 1992, J. Mol. Bio. 227:799-817. El directorio V BASE proporciona un directorio amplio de secuencias de región variable de inmunoglobulina humana (compilado por Tomlinson, I.A. et al. MRC Centre for Protein Engineering, Cambridge, UK). Los anticuerpos se "humanizan" (*germlined*) revirtiendo uno o más aminoácidos que no son de línea germinal en regiones del marco a los correspondientes aminoácidos de línea germinal del anticuerpo, siempre y cuando se retengan sustancialmente las propiedades de unión. Se pueden utilizar también métodos similares en la región constante, por ejemplo en dominios constantes de inmunoglobulina.

Por ejemplo, un anticuerpo puede incluir una, dos, tres, o más sustituciones de aminoácidos, por ejemplo, en un marco, CDR, o región constante, para que sea más similar a una secuencia de línea germinal de referencia. Un método ejemplar para la humanización (*germlining*) puede incluir identificar una o más secuencias de línea germinal que sean similares (por ejemplo, lo más similares en una base de datos en particular) a la secuencia del anticuerpo aislado. Entonces se realizan mutaciones (a nivel de los aminoácidos) en el anticuerpo aislado, ya sea en forma incremental o en combinación con otras mutaciones. Por ejemplo, se obtiene una biblioteca de ácidos nucleicos que incluye secuencias que codifican algunas o todas las mutaciones de línea germinal posibles. Los anticuerpos mutados se evalúan entonces, por ejemplo, para identificar un anticuerpo que tiene uno o más restos de línea germinal adicionales con respecto al anticuerpo aislado y que aún es útil (por ejemplo, tiene una actividad funcional). En una realización, se introducen tantos restos de línea germinal en un anticuerpo aislado como sea posible.

En una realización, se usa la mutagénesis para sustituir o insertar uno o más restos de línea germinal en un marco y/o región constante. Por ejemplo, un resto del marco y/o de región constante de línea germinal pueden ser de una secuencia de línea germinal que es similar (por ejemplo, muy similar) a la región no variable que está siendo modificada. Después de la mutagénesis, se puede evaluar la actividad (por ejemplo, la unión u otra actividad funcional) del anticuerpo para determinar si el resto o restos de línea germinal son tolerados (es decir, no eliminan la actividad). Se puede llevar a cabo una mutagénesis similar en las regiones del marco.

La selección de una secuencia de línea germinal se puede realizar de diferentes maneras. Por ejemplo, una secuencia de línea germinal se puede seleccionar si reúne unos criterios predeterminados de selectividad o similitud, por ejemplo al menos un cierto porcentaje de identidad, por ejemplo al menos 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, o 99,5% de identidad. La selección puede realizarse utilizando por lo menos 2, 3, 5, o 10 secuencias de línea germinal. En el caso de CDR1 y CDR2, la identificación de una secuencia de línea germinal similar puede incluir la selección de una de tales secuencias. En el caso de CDR3, la identificación de una secuencia de línea germinal similar puede incluir la selección de una de tales secuencias, pero puede incluir el uso de dos secuencias de línea germinal que contribuyen por separado a la porción amino terminal y la porción carboxi terminal. En otras implementaciones, se utilizan más de una o dos secuencias de línea germinal, por ejemplo, para formar una secuencia de consenso.

Diversidad de CDR1, CDR2, y de la cadena ligera

Se entiende que las bibliotecas de CDR3 de HC se construyen sobre la diversidad en CDR1 de HC, CDR2 de HC y cadenas ligeras. La diversidad de las cadenas ligeras se puede codificar en la misma molécula de ADN que la diversidad de HC, o las diversidades de LC y HC se pueden codificar en moléculas de ADN distintas. En la Tabla 22, la fusión de una secuencia señal::VH::CH1::His6::Myc::muñón III. CDR1 comprende los restos 31-35; hay diversidad en los restos 31, 33 y 35. En una realización, los restos 31, 33, y 35 pueden ser cualquier tipo de aminoácido excepto cisteína. CDR2 comprende los restos 50 a 65. Hay diversidad en las posiciones 50, 52, 52a, 56, y 58. En una realización, los restos 50 y 52 pueden ser cualquiera de los tipos Ser, Gly, Val, Trp, Arg, Tyr; el resto 52a puede ser Pro o Ser, y los restos 56 y 58 pueden ser cualquier tipo de aminoácido excepto Cys. La diversidad de CDR3 de HC se clona en una diversidad de CDR1 y 2 de HC que es al menos 1. E 4, 1. E 5, 1. E 6, 1. E 7, 5. E 7, ó 1. E 8.

En una realización, los restos 31, 33, 35, 50, 52, 56, y 58 pueden ser cualquier tipo de aminoácido excepto Cys o Met, y el resto 52a puede ser Gly, Ser, Pro, o Tyr. La diversidad de CDR3 de HC se clona en una diversidad de CDR1 y 2 de HC que es al menos 1. E 4, 1. E 5, 1. E 6, 1. E 7, 5. E 7, ó 1. E 8.

En una realización, la diversidad de la HC se clona en un vector (fago o fagómido) que contiene una diversidad de cadenas ligeras. Esta diversidad es al menos 25, 50, 100, 500, 1. E 3, 1. E 4, 1. E 5, 1. E 6, ó 1. E 7. La diversidad de CDR3 de HC es al menos 221, 272, 500, 1000, 1. E 4, 1. E 5, 1. E 6, 1. E 7, 1. E 8, ó 1. E 9.

En una realización, la diversidad de la HC se clona en un vector fágico que presenta la HC en una proteína de fago

tal como III, VIII, VII, VI, o IX, o un fragmento de uno de estos suficiente para provocar la presentación, y las cadenas ligeras se combinan con la HC infectando una colección de células en la que cada célula segrega una cadena ligera. La diversidad de las cadenas ligeras en las células es al menos 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, ó 100. La diversidad de CDR3 de HC es al menos 221, 272, 500, 1000, 1. E 4, 1. E 5, 1. E 6, 1. E 7, 1. E 8, ó 1. E 9.

- 5 La Tabla 30 muestra la secuencia del vector fágico DY3FHC87 (SEQ ID NO:894) que posee un gen *bla*, un casete de presentación para cadenas pesadas bajo el control de un promotor P_{lac} . DY3FHC87 contiene igualmente todos los genes de M13. Infectando células *E. coli* F+ que poseen una diversidad de cadenas ligeras en un vector tal como pLCSK23 (secuencia en la Tabla 40) (SEQ ID NO:896). El vector pLCSK23 posee el gen *Kan^R*. Bajo el control del promotor P_{lac} , hay un gen que comienza en la base 2215 que tiene una secuencia señal (bases 2215-2277), una VL (en esta secuencia la VL codifica la secuencia mostrada en (SEQ ID NO:897) desde la base 2278 hasta la base 2598, Ckappa desde la base 2599 a 2922, un ligador que permite un sitio NotI desde 2923 hasta 2931, y una etiqueta V5 (bases 2932-2973). Hay un sitio SfiI en 2259-2271 y un sitio *KpnI* en 2602-2605 para permitir la sustitución fácil de Vkappas. (SEQ ID NO:897) es un ejemplo de las proteínas que se segregan. Se ha de entender que CKappa y la etiqueta V5 son constantes. Todas las proteínas mostradas en la Tabla 19 (VK1O2gl-JK3, VK102var1, VK102var2, VK102var3, VK102var4, VK102var5, VK3L6gl-JK4, VK3L6var1, VK3L6var2, VK3L6var3, VK3L6var4, VK3L6var5, VK3L6var6, VK3L6var7, VK3L6var8, VK3A27gl-JK3, VK3A27var1, VK3A27var2, VK3A27var3, VK3A27var4, VK3A27var5, VK3A27var6, VK3A27var7, VK3L2gl-JK3, y VK1glL8-JK5) tendrán estas secuencias unidas en el extremo carboxi.

Diversidad de las cadenas ligeras

- 20 La Tabla 800 muestra una LC (cadena ligera) kappa que se sabe que se empareja bien con 3-23 y con cinco mutaciones de CDR con una HC basada en 3-23, LC K1(O12)::JK1 obtiene un Ab de alta afinidad por una diana proteica. 012 es un VKI usado frecuentemente. El gen se ha diseñado para tener sitios de restricción distintos, útiles, en la secuencia señal (*Apa*LI), FR1 (*Xho*I, *Sgf*I), FR2 (*Kpn*I), FR3 (*Xba*I), y Fr4::Ckappa (*Bs*WI), de manera que cada CDR y se reemplace por una población variada.

- 25 En LCs humanas, CDR3 es muy importante, y CDR1 es la siguiente en suma importancia. CDR2 raras veces entra en contacto con el Ag. La diversidad se introduce en las CDRs como se muestra en la Tabla 900 y la Tabla 1000 (CDR1), Tabla 1100 y Tabla 1200 (CDR2), Tablas 1300, 1400, y 1500 (CDR3). Para la selección económica de cadenas pesadas (ESHC), se escoge un número pequeño, por ejemplo 50 LCs con diversidad en CDR3 como en la Tabla 1200 para la expresión en pLCSK24 para la secreción en el periplasma. Se pueden usar más LCs si se mantienen varias estirpes celulares, de manera que cada estirpe celular contiene, por ejemplo, 50 o menor número de LC.

- 35 La Tabla 900 muestra diversidad para CDR1 de LC. La biblioteca puede contener el resto 012 con la diversidad añadida de los tipos AA mostrados como "permitidos"; leyéndose "permitidos" como "tipos permitidos adicionales" en las Tablas 900, 1000, 1100, 1200, 1300, 1400. 012 tiene R₂₄ASQSISSYLN₃₄. Otros loci VK1 tienen Q en 24. Otros loci tienen M en 25. S₂₆ y Q₂₇ son invariables en VKI. Otros loci VKI tienen D o G en 28. I₂₉ y L₃₃ son invariables en VKI, y los grupos laterales están orientados hacia dentro. Otros loci VKI permiten la diversidad mostrada en la Tabla 900 en las posiciones 30, 31, 32, y 34. En la Tabla 900, sólo se varían siete de las once posiciones, y la diversidad total es 576.

- 40 La Tabla 1000 muestra un mayor nivel de diversidad para CDR1 de LC. Aquí, se han variado 8 de 11 posiciones. Aquellas que son constantes están lejos del sitio de combinación o tienen grupos laterales escondidos.

La Tabla 1100 muestra una variegación de bajo nivel para CDR2. CDR2 está lejos del sitio de combinación del antígeno, y la diversidad aquí puede no ser muy útil. De hecho, la diversidad de GL está muy limitada. La Tabla 1100 incluye la diversidad de GL. La Tabla 1200 contiene un mayor nivel de diversidad, permitiéndose 1920 secuencias.

- 45 La Tabla 1300 muestra un bajo nivel de diversidad para CDR3 de LC, 2160 secuencias. La Tabla 1400 muestra un mayor nivel que permite 105.840 secuencias.

Para ROLIC, se producen alrededor de 3×10^7 LC que tienen la diversidad mostrada en las Tablas 900, 1100, y 1300.

Diversidad de la cadena pesada

- 50 Una HC (cadena pesada) de Ab tiene diversidad en CDR1, CDR2, y CDR3. La diversidad en CDR3 es especialmente compleja debido a que existe diversidad tanto de secuencia como de longitud. La diversidad de secuencia no es aleatoria. Las células que obtienen genes de Ab unen un segmento V a un segmento D a un segmento JH. El segmento D es opcional; alrededor de la mitad de los Abs humanos naturales tienen un D reconocible. Puede haber una edición amplia en las fronteras de V-D, D-J, o V-J, con ninguna a muchas bases añadidas o eliminadas. Un Ab que tiene un V::D::JH de línea germinal se debería ver como un Ab de línea germinal.

- 55 En la Tabla 21 se muestran los segmentos D humanos. Cada segmento D de línea germinal (GL) puede aparecer en un gen de Ab en cualquiera de los tres marcos de lectura directos. En algunos marcos de lectura, algunos de los

segmentos D codifican codones de parada. Estos segmentos D aparecen raramente con el codón de parada modificado. La Tabla 600 muestra la frecuencia de cada segmento D como un porcentaje de todos los segmentos D observados. La mayoría de los ejemplos aquí que contienen segmentos D usan Ds que son bastante habituales (> 2% de todos los Ds observados).

- 5 En un aspecto, la presente descripción implica componer genes de HC de Ab fusionando 3-23 (u otro VH, tal como 4-34) a uno de a) un número de aminoácidos recogidos del conjunto que comprende (S, Y, D, R, N), b) una región D, c) una región JH, y d) la porción FR4 de una región JH. Estas fusiones pueden ser un 3-23 de GL o un 3-23 que tiene diversidad sintética en CDR1 y/o CDR2. Las longitudes de la CDR3 de HC puede ser cualquier número de alrededor de 3 a alrededor de 24. Preferiblemente, la biblioteca contendría miembros con CDR3 de HC de longitudes
- 10 6, 8, 10, 12, 14, 16, 18, y 20. Como alternativa, las longitudes podrían ser 5, 8, 11, 14, 17, y 20, o cualquier otra combinación.

- La Tabla 21 muestra un número de ejemplos de diseños de CDR3s adecuadas con longitudes de 6 a 20. Los codones que especifican las letras mayúsculas en la columna 2 se sintetizan con titubeo. La columna 3 muestra el nivel de dopaje. La Tabla 100 muestra relaciones en las que las diversas longitudes de CDR3 de HC se podrían
- 15 combinar para formar una biblioteca que se espera que contenga Abs que se unen a casi todas las dianas proteicas.

Tabla 100							
Longitud	6	8	10	12	14	16	20
Diversidad	1×10^5	2×10^5	4×10^5	8×10^5	8×10^5	8×10^5	4×10^5

- Para la longitud 6, se dan cuatro ejemplos en la Tabla 21. Por ejemplo, 6a tiene VH(3-23) unido directamente a JH1 con los seis AAs titubeados, 6b tiene Tyr unido a D4-17 en segundo marco de lectura unido a los AAs de FR4 de JH1, y 6c tiene D5-5(3) unido a los restos FR de JH1. Puesto que estos dan diferentes tipos de diversidad, se
- 20 prefiere incluir todos, pero una biblioteca que contenga sólo uno de estos debería dar Abs útiles.

- Para la longitud 8, la Tabla 21 muestra tres ejemplos. 8a tiene YY fusionado a todos de JH1, mientras que 8b tiene un Y fusionado a D6-13(1) fusionado a la región FR de JH1. En la Tabla 21 también se muestran las longitudes 10, 12, 14, 16, y 20. La diversidad de CDR3 de HC se podría construir en un 3-23 de línea germinal o 3-23 que contiene diversidad sintética. Como alternativa, se podría usar un VH diferente, tal como 4-34.

- 25 ROLIC es un método en el que una pequeña población de HCs se expresa en F^+ *E. coli* como proteínas solubles. La población se infecta con fago que posee fusiones LC::III_{muñón}. El fago producido obtiene una HC del periplasma de la célula que las produce. Este fago se puede unir a una diana inmovilizada y el ligador se separa de los no ligadores. El tamaño de la población es importante debido a que cuando el fago recuperado se propaga, el fago recuperado debe encontrar el mismo tipo de célula de la que procede para continuar la asociación entre LC y HC. De este modo,
- 30 es deseable que el número de HC sea pequeño en cada estirpe celular. De este modo, puede ser deseable mantener un número de estirpes celulares con hasta 10, 20, 30, ó 40 HC diferentes en cada estirpe celular. De este modo, podemos tener 1, 2, 4, 6, 8, 10, 24, 48, ó 96 estirpes celulares y llevar a cabo el mismo número de producciones, selecciones y amplificaciones fágicas paralelas. Después de una o dos rondas, se ensayan colonias para la producción de fago que se une a la diana mediante un ensayo ELISA. Cada colonia de ELISA⁺ contiene una
- 35 LC útil y una HC útil, pero no están en el mismo trozo de ADN. No obstante, sabemos el comienzo y el final de cada LC y cada HC, y por lo tanto se puede usar PCR en la colonia para producir un casete de presentación de Fab o un casete de secreción de Fab que se puede colocar en un fago o fagómido de presentación o en un plásmido de producción de Fab.

- En la selección eficiente de HCs (ESHC), se invierten los papeles de LC y HC en ROLIC, y tenemos LCs en un plásmido de manera que se producen como proteínas solubles en el periplasma de F^+ *E. coli*. Se produce la diversidad de HC en un vector fágico que no tiene gen de LC. Se infecta la F^+ *E. coli* productora de LC con el fago que porta HC. Se obtiene un fago que porta un gen de HC y proteínas tanto de HC como de LC. Se seleccionan estos fagos para la unión a la diana. En muchos Abs, la LC es permisiva y no contribuye enormemente a la afinidad de unión. Tomando la mejor LC se puede incrementar enormemente la afinidad, pero habitualmente es posible
- 40 seleccionar un Fab con un repertorio muy limitado de LCs. De este modo, se coloca un pequeño conjunto de LCs, preferiblemente línea germinal en las regiones del marco en la F^+ *E. coli* productora de LC. Si, por ejemplo, hay 25 LC en la estirpe celular de LC, entonces se obtiene una reducción de 25 veces en el número de transformantes celulares que se necesitan obtener.

- Las bibliotecas descritas tienen un intervalo de longitudes de CDR3 de HC. Para favorecer el plegamiento apropiado, la CDR3 de HC tiene un segmento D o ningún segmento D unido a casi todo, a todo, o a la porción del marco de un segmento JH. Las secuencias se diversifican usando síntesis de ADN de titubeo. Aunque esto permite
- 50 teóricamente cualquier tipo de aminoácido en cualquier posición, en la práctica las secuencias reales están fuertemente desviadas con respecto a las secuencias progenitoras y tipos de AA que están próximos en la tabla del código genético.

Usando ESHC, se pueden muestrear nuevos diseños de diversidad de CDR3 de HC sintética. En los ejemplos dados, se usa un conjunto de, por ejemplo, 50 LCs. Una biblioteca de 5×10^8 HC debería comportarse igual de bien que una biblioteca del estilo antiguo de $2,5 \times 10^{10}$, pero requiere bastante menos esfuerzo.

- 5 Cuando se titubea una secuencia, la recogida de los codones iniciales afecta a la mezcla real de AAs observada en la biblioteca. La Tabla 300 muestra qué sustituciones de aminoácidos requieren 1, 2 ó 3 cambios de bases a partir de cada codón progenitor de partida. Por ejemplo, si comenzamos con gct o gcc para Ala, los tres codones de parada requieren tres cambios de bases, y de este modo son raros. Si se usan mezclas 76:8:8:8, Ala aparecerá en el 57% de los casos ($0,76 \times 0,76$). V, G, T, P, S aparecerán cada uno en alrededor de 6%, y D alrededor de 3%. E, I, L, F, Y, H, N, C, y R estarán por debajo de alrededor de 10 veces. M, W, Q, K, Am, Oc, y Op serán incluso más raros.
- 10 Si comenzamos con gca, entonces E sustituirá a D a la hora de necesitar solamente un cambio de base, pero las paradas opal y ochre requieren sólo dos cambios de bases, lo que es indeseable. Los codones preferidos están marcados con una estrella (*). La elección para serina es complicada; nuestro deseo es sustituir Y por S con frecuencia elevada. Esto trae a Op y Oc en el grupo que difiere del progenitor en solamente dos bases. Este problema se puede resolver clonando el repertorio de CDR3 de HC antes de un gen de resistencia a antibiótico, tal como KanR o AmpR y seleccionando en busca de resistencia, eliminando así los miembros que contienen codones de parada.
- 15 Además, la biblioteca se puede producir en supE *E. coli* que inserta Q en lugar de la parada.

Tabla 300

Aminoácido	Codón progenitor	1 cambio de base	2 cambios de base	3 cambios de base
A *	gct, gcc	V, D, G, T, P, S	E, I, L, F, Y, H, N, C, R	M, W, Q, K, Am, Oc, Op
A	gca	V, E, G, T, P, S	D, I, L, Oc, Q, K, Op, R	M, W, H, N, C, Am, F, Y
A	gcg	V, E, G, T, P, S	D, M, L, Am, Q, K, R, W	I, F, Y, Oc, Op, H, N, C
C	tgt, tgc	Y, S, F, W, Op, R, G	L, H, N, D, P, T, A, V, I	Am, Oc, Q, K, E, M
D	gat, gac	E, G, A, V, N, H, Y	F, S, C, L, P, Q, K, R, Oc, Am, I, T	M, W, Op
E	gaa	D, G, A, V, K, Q, Oc	Am, L, I, S, P, T, R, Op, Y, H, N	M, F, C, W
E *	gag	D, G, A, V, K, Q, Am	M, L, S, P, T, Y, H, N, Oc, R, W	F, C, I, Op
F	ttt, ttc	L, I, V, S, Y, C	M, Am, Op, Oc, W, P, T, A, H, N, D, R, G	Q, K, E
G *	ggg, ggc	D, A, V, S, R, C	E, W, F, L, I, T, P, Y, H, N	Am, Oc, Op, M, Q, K
G	gga	E, A, V, R, Oc	D, W, L, I, S, P, T, Op, Q, K	Am, Oc, M, F, Y, H, N
G	ggg	E, A, V, R, W	D, Oc, L, M, S, P, T, Am, Op, Q, K	Oc, I, F, Y, H, N
H	cat, cac	Q, Y, N, D, L, P, R	F, S, C, I, T, V, A, D, G, Am, Oc	Op, W, M, E
I *	att, atc	M, L, F, V, T, N, S	Y, C, P, H, R, A, D, G	Am, Op, Oc, W, Q, K, E
I	ata	M, L, V, T, K, R	Op, Oc, S, P, Q, A, E, G, F, N	Am, C, D, H, W, Y
K	aaa	N, Q, Oc, E, P, I, R	H, Y, D, M, L, V, S, T, A, Am, Op, G	C, F, W
K *	aag	N, Q, Am, E, P, M, R	H, Y, D, I, L, V, S, T, A, Oc, G, W	C, F, Op
L	tta	F, S, Oc, Op, I, V	Y, C, W, M, P, T, A, Q, K, E, R, G, Am	D, H, N
L	ttg	F, S, Am, W, M, V	Y, C, Oc, Op, P, T, A, Q, K,	D, H, N

ES 2 528 963 T3

Aminoácido	Codón progenitor	1 cambio de base	2 cambios de base	3 cambios de base
			E, R, G, I	
L *	ctt, ctc	F, I, V, P, H, R	M, S, Y, C, T, N, A, D, G	Am, Oc, Op, W, E, K, Q
L	cta	I, V, P, Q, R	F, M, S, Oc, Op, T, K, A, E, G, H	Am, W, D, N, C, Y
L	ctg	M, V, P, Q, R	F, I, S, Am, T, K, A, E, G, H, W	Oc, Op, D, N, C, Y
M	atg	L, V, T, K, R, I	F, N, S, P, A, Am, Q, E, W, G	Oc, Op, Y, C, H, D
N	aat, aac	K, Y, H, D, I, T, S	F, C, L, P, R, V, A, G, M, Q, E, Am, Oc	Op, W
P*	cct,ccc	S, T, A, L, H, R	F, Y, C, I, N, V, D, G, Q	Am, Oc, Op, W, M, E, K
P	cca	S, T, A, L, Q, R	Oc, Op, I, K, V, E, G, H	Am, W, M, D, N, C, F, Y
P	ccg	S, T, A, L, Q, R	Am, M, K, V, E, G, H	C, D, F, I, N, W, Y, Oc, Op
Q	caa	Oc, K, E, R, P, L, H	Y, Am, N, D, S, T, A, I, V, G, Op	F, C, W, M
Q*	cag	H, Am, K, E, R, P, L	N, D, Y, M, T, V, A, G, W, Oc, S	C, F, Op, I
R*	cgt, cgc	C, S, G, H, P, L	Op, W, Q, F, Y, I, T, N, V, A, D	Am, Oc, M, E, K
R	cga	G, Op, Q, P, L	Oc, S, C, W, H, I, V, T, A, E, K	Am, M, C, D, N, F, Y
R	cgg	G, W, Q, P, L	Am, Op, S, M, V, T, A, K, E, H, C	F, Y, I, Oc, D, N
R	aga	G, Op, S, K, T, I	C, W, N, M, L, V, P, A, Oc, Q, E	F, Y, H, D, Am
R	agg	G, W, S, K, T, M	C, Op, Am, L, I, V, A, Q, P, E, N	F, Y, H, D, Oc
S *	tct, tcc	F, Y, C, P, T, A	L, Oc, Op, Am, W, I, V, N, D, R, G, H	E, K, M, Q
S	tca	L, Oc, Op, P, T, A	F, Y, C, W, Q, R, I, K, V, E, G, Am	M, W, D, N, H
S	tcg	L, Am, W, P, T, A	F, Y, C, Op, Oc, Q, R, M, K, V, E, G	I, D, N, H
S	agt, agc	C, R, G, N, T, I	F, Y, L, P, H, V, A, D, K, W, Op	Am, Oc, M, E, Q
T*	act, acc	S, P, A, I, N	F, Y, C, L, H, R, M, K, V, D, G	Am, Oc, Op, W, E, Q
T	aca	S, P, A, I, K, R	L, Oc, Op, Q, M, E, G, V, N	F, Y, C, Am, W, D, H
T	acg	S, P, A, M, K, R	I, N, L, Am, W, Q, V, E, G	C, F, Y, Oc, Op, D, H
V*	gtt, gtc	F, L, I, A, D, G	S, P, T, Y, H, N, E, C, R, M	Am, Oc, Op, W, Q, K
V	gta	L, I, A, E, G	F, M, D, S, P, T, Oc, Op, Q	Am, W, C, Y, H, N

Aminoácido	Codón progenitor	1 cambio de base	2 cambios de base	3 cambios de base
			R, K	
V	gtg	L, M, A, E, G	F, I, D, S, P, T, Am, Q, R, K, W	Oc, Op, C, Y, H, N
W	tgg	C, R, G, Am, S, L, Op	P, Q, F, M, T, K, V, A, E, Oc, Y	D, N, H, I
Y	tat, tac	C, S, F, N, H, D, Oc, Am	L, W, Q, K, E, P, I, T, V, A, G, Op, R	M
Am es parada TAG, Op es TGA, Oc es TAA				

Método para usar las bibliotecas

Selección de la constante de disociación. Dado que una velocidad de disociación lenta puede ser predictiva de una alta afinidad, en particular con respecto a las interacciones entre polipéptidos y sus dianas, los métodos descritos aquí pueden ser utilizados para aislar ligandos con una velocidad de disociación cinética deseada (por ejemplo, reducida) para una interacción de unión con una diana.

Para seleccionar anticuerpos que se disocian lentamente a partir de una biblioteca de presentación, la biblioteca se pone en contacto con una diana inmovilizada. La diana inmovilizada se lava luego con una primera disolución que elimina los anticuerpos unidos de forma no específica o débilmente. Después los anticuerpos unidos se eluyen con una segunda disolución que incluye una cantidad saturante de diana libre, es decir, réplicas de la diana que no se unen a la partícula. La diana libre se une a anticuerpos que se disocian de la diana. La reunión de los anticuerpos eluidos se evita efectivamente mediante la cantidad saturante de diana libre con respecto a la concentración mucho menor de la diana inmovilizada.

La segunda disolución puede tener condiciones de disolución que son sustancialmente fisiológicas o que son rigurosas (por ejemplo, pH bajo, pH alto, o alto contenido de sal). Típicamente, las condiciones de disolución de la segunda disolución son idénticas a las condiciones de disolución de la primera disolución. Las fracciones de la segunda disolución se recogen en orden temporal para distinguir las primeras fracciones de las fracciones finales. Las fracciones finales incluyen anticuerpos que se disocian a una velocidad más lenta de la diana que las biomoléculas en las primeras fracciones. Además, también es posible recuperar anticuerpos que permanecen unidos a la diana, incluso después de una incubación prolongada. Estos pueden disociarse utilizando condiciones caotrópicas o pueden amplificarse mientras están unidos a la diana. Por ejemplo, se puede poner en contacto el fago unido a la diana con las células bacterianas.

Selección o identificación para especificidad. Los métodos de identificación de bibliotecas de presentación descritos en este documento pueden incluir un procedimiento de selección o identificación que descarta anticuerpos que se unen a una molécula no diana. Los ejemplos de moléculas no diana incluyen, por ejemplo, una molécula de hidrato de carbono que difiere estructuralmente de la molécula diana, por ejemplo una molécula de hidrato de carbono que tiene una propiedad biológica diferente de la molécula diana. En el caso de un hidrato de carbono sulfatado, una no diana puede ser el mismo hidrato de carbono sin el sulfato o con el sulfato en una posición diferente. En el caso de un fosfopéptido, la no diana puede ser el mismo péptido sin el fosfato o un fosfopéptido diferente.

En una implementación, se utiliza una etapa denominada de "selección negativa" para discriminar entre la diana y la molécula no diana relacionada y moléculas no diana relacionadas pero distintas. La biblioteca de presentación o un conjunto de la misma se pone en contacto con la molécula no diana. Los miembros que no se unen a la no diana se recogen y se utilizan en selecciones posteriores para la unión a la molécula diana o incluso para selecciones negativas posteriores. La etapa de selección negativa puede ser antes o después de seleccionar miembros de la biblioteca que se unen a la molécula diana.

En otra implementación, se utiliza una etapa de identificación. Después de aislar los miembros de la biblioteca de presentación por la unión a la molécula diana, cada miembro aislado de la biblioteca se ensaya en busca de su capacidad para unirse a una molécula no diana (por ejemplo, una no diana listada más arriba). Por ejemplo, se puede utilizar una identificación por ELISA de alto rendimiento para obtener este dato. También se puede utilizar la identificación por ELISA para obtener datos cuantitativos para la unión de cada miembro de la biblioteca a la diana. Los datos de unión con la diana y la no diana se comparan (por ejemplo, mediante el uso de un ordenador y un software) para identificar miembros de la biblioteca que se unen específicamente a la diana.

En ciertas realizaciones, los anticuerpos que comprenden las CDR3s de la invención pueden ser capaces de unirse a hidratos de carbono. Los métodos para evaluar anticuerpos para la unión a hidratos de carbono incluyen ELISA,

inmunohistoquímica, inmunotransferencia, y clasificación celular activada por fluorescencia. Estos métodos se pueden usar para identificar anticuerpos que tienen una K_D mejor que un umbral, por ejemplo mejor que 100 nM, 50 nM, 10 nM, 5 nM, 1 nM, 500 pM, 100 pM, o 10 pM.

5 ELISA. Las proteínas codificadas por una biblioteca de presentación también se pueden identificar en busca de una propiedad de unión usando un ensayo ELISA. Por ejemplo, cada proteína se pone en contacto con una placa de microtitulación cuya superficie del fondo se ha recubierto con la diana, por ejemplo una cantidad limitante de la diana. La placa se lava con tampón para eliminar polipéptidos unidos no específicamente. Después, la cantidad de la proteína unida a la placa se determina sondando la placa con un anticuerpo que puede reconocer el polipéptido, por ejemplo una etiqueta o porción constante del polipéptido. El anticuerpo se enlaza a una enzima, tal como fosfatasa alcalina, que produce un producto colorimétrico cuando se proporcionan sustratos apropiados. La proteína se puede purificar a partir de células o se puede evaluar en un formato de biblioteca de presentación, por ejemplo como una fusión a un recubrimiento de bacteriófago filamentosos. Como alternativa, las células (por ejemplo, vivas o fijas) que expresan la molécula diana, por ejemplo una diana que contiene un resto de hidrato de carbono, se pueden colocar en una placa de microtitulación y se pueden usar para ensayar la afinidad de los péptidos/anticuerpos presentes en la biblioteca de presentación, o se pueden obtener mediante selección de la biblioteca de presentación.

En otra versión del ensayo ELISA, cada polipéptido de una biblioteca de hebras de diversidad se usa para recubrir un pocillo diferente de una placa de microtitulación. Entonces el ELISA transcurre usando una molécula diana constante para investigar cada pocillo.

20 Ensayos de unión a células. Los anticuerpos se pueden evaluar para determinar su capacidad para interactuar con uno o más tipos celulares, por ejemplo una célula hematopoyética. La clasificación celular activada por fluorescencia (FACS) es un método ejemplar para ensayar una interacción entre una proteína y una célula. En anticuerpo se marca directa o indirectamente con un fluoróforo, antes o después de la unión a las células, y después las células se cuentan en un clasificador de FACS.

Se pueden preparar otros tipos celulares para FACS mediante métodos conocidos en la técnica.

25 Ensayos de unión homogéneo. La interacción de unión de un polipéptido candidato con una diana puede ser analizada utilizando un ensayo homogéneo, es decir, después de añadir todos los componentes del ensayo, no se requieren manipulaciones adicionales de fluidos. Por ejemplo, se puede usar transferencia de energía de resonancia de fluorescencia (FRET) como un ensayo homogéneo (véase, por ejemplo, Lakowicz et al., patente U.S. nº 5.631.169; Stavrianopoulos, et al., patente U.S. nº 4.868.103). Se selecciona un marcador fluoróforo en la primera molécula (por ejemplo, la molécula identificada en la fracción) de tal manera que su energía fluorescente emitida puede ser absorbida por un marcador fluorescente en una segunda molécula (por ejemplo, la diana) si la segunda molécula se encuentra próxima a la primera molécula. El marcador fluorescente en la segunda molécula emite fluorescencia cuando absorbe a la energía transferida. Dado que la eficiencia de transferencia de energía entre los marcadores está relacionada con la distancia que separa las moléculas, se puede evaluar la relación espacial entre las moléculas. En una situación en la que se produce unión entre las moléculas, la emisión fluorescente del marcador de la molécula "aceptora" en el ensayo debe ser máxima. Un suceso de unión que está configurado para la monitorización mediante FRET se puede medir convenientemente a través de un medio de detección fluorométrico estándar bien conocido en la técnica (por ejemplo, usando un fluorímetro). Mediante la titulación de la cantidad de la primera o la segunda molécula de unión, se puede generar una curva de unión para estimar la constante de unión del equilibrio.

Otro ejemplo de un ensayo homogéneo es Alpha Screen (Packard Bioscience, Meriden Conn.). Alpha Screen utiliza dos perlas marcadas. Una perla genera oxígeno singlete cuando es excitado por un láser. La otra perla genera una señal luminosa cuando el oxígeno singlete se difunde desde la primera perla y choca con ella. La señal sólo se genera cuando las dos perlas están próximas. Una perla puede estar unida al miembro de la biblioteca de presentación, la otra a la diana. Las señales se miden para determinar el grado de unión.

Los ensayos homogéneos se pueden llevar a cabo mientras que el polipéptido candidato esté unido al vehículo de la biblioteca de presentación, por ejemplo un bacteriófago.

50 Resonancia de plasmones de superficie (SPR). La interacción de unión de una molécula aislada de una biblioteca de presentación y una diana se puede analizar usando SPR. La SPR o el análisis de interacción biomolecular (BIA) detectan interacciones bioespecíficas en tiempo real, sin marcar ninguno de los interactuantes. Los cambios en la masa en la superficie de unión (indicativos de un suceso de unión) del chip de BIA dan como resultado alteraciones del índice de refracción de la luz cerca de la superficie (el fenómeno óptico de resonancia de plasmones de superficie (SPR)). Los cambios en la refractividad generan una señal detectable, que se mide como una indicación de las reacciones en tiempo real entre moléculas biológicas. Los métodos para utilizar SPR se describen, por ejemplo, en la patente U.S. nº 5.641.640; Raether, 1988, Surface Plasmons Springer Verlag; Sjolander y Urbaniczky, 1991, Anal. Chem. 63: 2338-2345; Szabo et al., 1995, Curr. Opin. Struct. Biol. 5: 699-705 y recursos en línea proporcionados por BIAcore International AB (Uppsala, Suecia).

La información de la SPR se puede utilizar para proporcionar una medida precisa y cuantitativa de la constante de disociación en el equilibrio (K_D), y parámetros cinéticos, incluyendo $k_{\text{asociación}}$ y $k_{\text{disociación}}$, para la unión de una biomolécula a una diana. Estos datos pueden ser utilizados para comparar diferentes biomoléculas. Por ejemplo, las proteínas codificadas por ácido nucleico seleccionado de una biblioteca de hebras de diversidad se pueden comparar para identificar individuos que tienen alta afinidad por la diana, o que tienen una $k_{\text{disociación}}$ lenta. Esta información también puede ser utilizada para desarrollar las relaciones estructura-actividad (SAR). Por ejemplo, los parámetros cinéticos y de unión en el equilibrio de las versiones maduras de una proteína progenitora se pueden comparar con los parámetros de la proteína progenitora. Se pueden identificar aminoácidos variantes en posiciones dadas que se correlacionan con determinados parámetros de unión, por ejemplo una alta afinidad y $k_{\text{disociación}}$ lenta. Esta información puede combinarse con el modelamiento estructural (por ejemplo, usando modelamiento por homología, minimización de la energía, o determinación de la estructura por cristalografía de rayos X o RMN). Como resultado, se puede formular una comprensión de la interacción física entre la proteína y su diana, y usarla para guiar otros procesos de diseño.

Matrices de proteínas. Las proteínas identificadas de la biblioteca de presentación se pueden inmovilizar sobre un soporte sólido, por ejemplo en una perla o en una matriz. Para una matriz de proteína, cada uno de los polipéptidos se inmoviliza en una localización única en un soporte. Típicamente, la localización es una localización bidimensional. Los métodos para producir matrices de polipéptidos se describen, por ejemplo, en De Wildt et al. (2000) Nat. Biotechnol. 18: 989-994; Lueking et al. (1999) Anal. Biochem. 270: 103-111; Ge (2000) Nucleic Acids Res. 28, e3, I-VII; MacBeath y Schreiber (2000) Science 289: 1760-1763; los documentos WO 01/40803 y WO 99 / 51773A1. Los polipéptidos para la matriz se pueden observar a gran velocidad, por ejemplo usando aparatos robóticos comercialmente disponibles, por ejemplo de Genetic MicroSystems o BioRobotics. El sustrato de la matriz puede ser, por ejemplo, nitrocelulosa, plástico, vidrio, por ejemplo vidrio modificado en la superficie. La matriz también puede incluir una matriz porosa, por ejemplo acrilamida, agarosa, u otro polímero.

Kits

También se proporcionan kits para uso para llevar a cabo un método según cualquier aspecto de la invención. Los kits pueden incluir los vectores necesarios. Uno de tales vectores tendrá típicamente un origen de replicación para un bacteriófago monocatenario, y contiene el ácido nucleico del miembro spb o tiene un sitio de restricción para su inserción en la región del extremo 5' de la secuencia codificante madura de una proteína de la cápside del fago, y con una secuencia codificante directriz secretora en dirección 5' de dicho sitio, que dirige una fusión del polipéptido exógeno de la proteína de la cápside hacia el espacio periplásmico.

También se proporcionan paquetes que codifican las CDR3s de HC como se definen anteriormente, y polipéptidos que comprenden las CDR3s de HC y sus fragmentos y derivados, obtenibles mediante uso de cualquiera de los métodos definidos anteriormente. Los derivados pueden comprender polipéptidos fusionados a otra molécula tal como una enzima o una cola Fc.

El kit puede incluir un vector fágico (por ejemplo DY3F87HC) que tiene un sitio para inserción de CDR3s de HC para la expresión del polipéptido codificado en forma libre. El kit también puede incluir un vector plasmídico para la expresión de cadenas ligeras solubles, por ejemplo pLCSK23. El kit también puede incluir una estirpe celular adecuada (por ejemplo, TG1). La diversidad de cadenas ligeras codificadas por pLCSK23 puede ser 10, 15, 20, 25, 30, ó 50. Las LCs en la diversidad se pueden construir o recoger para que tengan ciertas propiedades deseables, tal como ser la línea germinal en las regiones del marco y tener diversidad en CDR3 y/o CDR1. Las líneas germinales pueden ser las más utilizadas, por ejemplo VK1_2-02, VK3_1-A27, VK3_5-L6, VK3_3-L2 para kappa y VL2_2a2, VL1_1c, VL1_1g, VL3_3r para lambda.

Por ejemplo, se podrían clonar genes para VK1O2g1-JK3, VK1O2var1, VK1O2var2, VK1O2var3, VK1O2var4, VK1O2var5, VK3L6g1-JK4, VK3L6var1, VK3L6var2, VK3L6var3, VK3L6var4, VK3L6var5, VK3L6var6, VK3L6var7, VK3L6var8, VK3A27g1-JK3, VK3A27var1, VK3A27var2, VK3A27var3, VK3A27var4, VK3A27var5, VK3A27var6, VK3A27var7, VK3L2g1-JK3, VK1g1L8-JK5, y VK1GLO12-JK3 (secuencias de aminoácidos mostradas en la Tabla 19) en pLCSK23.

Tabla 19: 26 VL a usar en pLSK23.

ES 2 528 963 T3

```

VK102ql-JK3                      (SEQ ID NO:4)
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQKP GKAPKLLIYA ASSLQSGVPS 60
RFGSGSGGTD FTLTISSLQP EDFATYYCQQ SYSTPFTFGP GTKVDIK 107

VK102var1                      (SEQ ID NO:5) S28D
DIQMTQSPSS LSASVGDRVT ITCRASQDIS SYLNWYQQKP GKAPKLLIYA ASSLQSGVPS 60
RFGSGSGGTD FTLTISSLQP EDFATYYCQQ SYSTPFTFGP GTKVDIK 107

VK102var2                      (SEQ ID NO:6) S91R
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQKP GKAPKLLIYA ASSLQSGVPS 60
RFGSGSGGTD FTLTISSLQP EDFATYYCQQ RYSTPFTFGP GTKVDIK 107

VK102var3                      (SEQ ID NO:7) S91E
DIQMTQSPSS LSASVGDRVT ITCRASQSI SYLNWYQQKP GKAPKLLIYA ASSLQSGVPS 60
RFGSGSGGTD FTLTISSLQP EDFATYYCQQ EYSTPFTFGP GTKVDIK 107

VK102var4                      (SEQ ID NO:8) S31R
DIQMTQSPSS LSASVGDRVT ITCRASQSI RYLNWYQQKP GKAPKLLIYA ASSLQSGVPS 60
RFGSGSGGTD FTLTISSLQP EDFATYYCQQ SYSTPFTFGP GTKVDIK 107

VK102var5                      (SEQ ID NO:9) S31E, S93R
DIQMTQSPSS LSASVGDRVT ITCRASQSI EYLNWYQQKP GKAPKLLIYA ASSLQSGVPS 60
RFGSGSGGTD FTLTISSLQP EDFATYYCQQ SYRTPFTFGP GTKVDIK 107

VK3L6ql-JK4                      (SEQ ID NO:10)
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQQKP GQAPRLLIYD ASNRATGIPA 60
RFGSGSGGTD FTLTISSLEP EDFAVYYCQQ RSNWPLTFGG GTKVEIK 107

VK3L6var1                      (SEQ ID NO:11) S31R
EIVLTQSPAT LSLSPGERAT LSCRASQSVS RYLAWYQQKP GQAPRLLIYD ASNRATGIPA 60
RFGSGSGGTD FTLTISSLEP EDFAVYYCQQ RSNWPLTFGG GTKVEIK 107

VK3L6var2                      (SEQ ID NO:12) S92R

```

EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RRNWPLTFGG GTKVEIK	107
VK3L6var3 (SEQ ID NO:13) S92G	
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RGNWPLTFGG GTKVEIK	107
VK3L6var4 (SEQ ID NO:14) S92Y	
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RYNWPLTFGG GTKVEIK	107
VK3L6var5 (SEQ ID NO:15) S92E	
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RENWPLTFGG GTKVEIK	107
VK3L6var6 (SEQ ID NO:16) Y32F	
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SFLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RSNWPLTFGG GTKVEIK	107
VK3L6var7 (SEQ ID NO:17) Y32D	
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SDLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RSNWPLTFGG GTKVEIK	107
VK3L6var8 (SEQ ID NO:18) N93G	
EIVLTQSPAT LSLSPGERAT LSCRASQSVS SYLAWYQQKP GQAPRLLIYD ASNRATGIPA	60
RFSGSGSGTD FTLTISSLEP EDFAVYYCQQ RSGWPLTFGG GTKVEIK	107
VK3A27q1-JK3 (SEQ ID NO:19)	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSSPFTFG PGTKVDIK	108
VK3A27var1 (SEQ ID NO:20) S31R	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS RSYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSSPFTFG PGTKVDIK	108
VK3A27var2 (SEQ ID NO:21) S32R	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SRYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSSPFTFG PGTKVDIK	108
VK3A27var3 (SEQ ID NO:22) S32D	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SDYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGSSPFTFG PGTKVDIK	108
VK3A27var4 (SEQ ID NO:23) G93E	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYESSPFTFG PGTKVDIK	108
VK3A27var5 (SEQ ID NO:24) G93R	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYRSPFTFG PGTKVDIK	108
VK3A27var6 (SEQ ID NO:25) S30D, G93E	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYESSPFTFG PGTKVDIK	108
VK3A27var7 (SEQ ID NO:26) S94R	
EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSYLAWYQQK PGQAPRLLIY GASSRATGIP	60
DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QYGRSPFTFG PGTKVDIK	108
VK3L2q1-JK3 (SEQ ID NO:27)	
EIVMTQSPAT LSVSPGERAT LSCRASQSVS SNLAWYQQKP GQAPRLLIYG ASTRATGIPA	60
RFSGSGSGTE FTLTISSLQ EDFAVYYCQQ YNNWPFITFGP GTKVDIK	107
VK1q1L8-JK5 (SEQ ID NO:28)	
DIQLTQSPSF LSASVGDRVT ITCRASQGIS SYLAWYQQKP GKAPKLLIYA ASTLQSGVPS	60
RFSGSGSGTE FTLTISSLQ EDFAVYYCQQ LNSYPITFGG GTRLEIK	107
VK1GLO12-JK3 (SEQ ID NO:897)	
DIQMTQSPSS LSASVGDRV TITCRASQSI SSYLNWYQQK PGKAPKLLIY AASSLQSGVP	60
SRFSGSGSGT DFTLTISL QPEDFATYYC QQSSTPFTF GPGTKVDIKR GTVAAPSVFI	120
FPPSDEQLKS GTASVVCLL NNFYPREAKV QWKVDNALQS GNSQESVTEQ DSKDSTYSL	180
STLILSKADY EKHKVYACE VTHQGLSSPV TKSFNRGCA AAGKPIPNPL LGLDST	236

Los kits pueden incluir componentes auxiliares requeridos para llevar a cabo el método, por supuesto dependiendo la naturaleza de tales componentes del método particular empleado. Los componentes auxiliares útiles pueden comprender fago auxiliar, cebadores de PCR, tampones, y/o enzimas de diversos tipos. Los tampones y enzimas se

usan típicamente para permitir la preparación de secuencias nucleotídicas que codifican fragmentos Fv, scFv o Fab derivados de genes de inmunoglobulinas reorganizados o no reorganizados según las estrategias descritas aquí.

MÉTODOS PARA INTRODUCIR DIVERSIDAD

Hay muchas maneras para generar ADN que sea variable. Una manera es usar síntesis nucleotídica mixta (MNS).

- 5 Una versión de MNS usa mezclas equimolares de nucleótidos como se muestra en la Tabla 5. Por ejemplo, el uso de codones NNK da los veinte aminoácidos y un codón de parada TAG. La distribución es 3(R/S/L): 2(A/G/V/T/P):

10 1(C/D/E/F/H/I/K/M/N/Q/W/Y) (por ejemplo, 3 de cada uno de Arg, Ser, y Leu, y así sucesivamente). Una alternativa, denominada aquí "titubeo", usa nucleótidos mixtos pero no en cantidades equimolares. Por ejemplo, si un codón progenitor fuese TTC (que codifica Phe), se podría usar una mezcla de (0,082 T, 0,06 C, 0,06 A, y 0,06 G) en lugar de T, y una mezcla de (0,082 C, 0,06 T, 0,06 A, y 0,06 G) en lugar de C. Esto podría dar TTC o TTT (que codifica Phe) 59% de las veces, y Leu 13%, S/V/I/C/Y ~5%, y otros tipos de aminoácidos menos a menudo.

15 Van den Brulle et al. (Biotechniques 45:340-3 (2008)) describen un método de síntesis de ADN variable en el que se usan enzimas de restricción tipo IIs para transferir trinucleótidos desde un oligonucleótido de horquilla de pelo anclado (PHONs) a un denominado "esplínquer". Véanse también las patentes EP 1.181.395, EP 1.411.122, EP 1.314.783 y las solicitudes EP 01127864.5, EP 04001462.3, EP 08006472.8. Usando mezclas de PHONs anclados y esplínquers, se pueden construir bibliotecas en las que se permiten tipos de aminoácidos deseados en relaciones determinadas por el diseñador. De este modo, se puede hacer que esté presente un tipo de aminoácido, por ejemplo 82% de las veces, y 18 de los otros tipos de aminoácidos (todos los tipos de aminoácidos no progenitores excepto Cys) estén presentes cada uno un 2%. Aquí, nos referiremos a tal síntesis como "titubeo digital" (titubeo digital). En algunos aspectos, el titubeo digital se prefiere al titubeo, pero el titubeo proporciona realizaciones útiles, parcialmente debido a que la estructura de la tabla del código genético hace que el titubeo realice sustituciones mayoritariamente conservativas. El titubeo digital ofrece la posibilidad de excluir tipos de aminoácidos indeseados. En las CDRs, se conocen cisteínas no emparejadas, incluso en Abs aprobados como sustancias terapéuticas, pero, en algunas realizaciones, se desearía evitarlas. En algunas realizaciones, cuando se diversifica una región D que contiene un par de cisteínas, no se permite que las cisteínas varíen debido a que el bucle cerrado de disulfuro es un elemento estructural importante y debido a que no se desea cisteínas no emparejadas.

25 Además, se puede sintetizar una molécula de ADN que codifica una secuencia de aminoácidos progenitora, y someter a ese ADN a PCR con tendencia a error usando cebadores que cubren las regiones del marco de manera que se eviten mutaciones en las regiones del marco.

Tabla 5: Códigos estándar para nucleótidos mixtos

N es A, C, G, T equimolares	
B es C, G, T equimolares	(no A)
D es A, G, T equimolares	(no C)
H es A, C, T equimolares	(no G)
V es A, C, G equimolares	(no T)
K es G, T equimolares	(Keto = ceto)
M es A, C equimolares	(aMino)
R es A, G equimolares	(puRina)
S es C, G equimolares	(Strong = fuerte)
W es A, T equimolares	(weak = débil)
Y es C, T	(pYrimidine = pirimidina)

Tabla 6: Ejemplo de nucleótidos mixtos para titubeo

e = 0,82 A + 0,06 C + 0,06 G + 0,06 T
q = 0,06 A + 0,82 C + 0,06 G + 0,06 T

$$j = 0,06 A + 0,06 C + 0,82 G + 0,06 T$$

$$z = 0,06 A + 0,06 C + 0,06 G + 0,82 T$$

EJEMPLIFICACIÓN

La presente invención se ilustra adicionalmente mediante los siguientes ejemplos, que de ninguna manera se deben de interpretar como limitantes.

5 Ejemplo profético 1: Bibliotecas con CDR3s de HC muy cortas

En la técnica se han descrito CDR3s de HC muy cortas. Kadirvelraj et al. (2006) Proc. Natl. Acad. Sci. USA 103:8149-54 han descrito una secuencia de CDR3 de HC de cuatro aminoácidos en un anticuerpo que se une a Ag de *Streptococcus* Tipo B III (GBS-Ag) pero no a Ag capsular de *Streptococcus pneumoniae*. GBS-Ag está sialilado a intervalos regulares. Ag capsular de *S. pneumoniae* (SPC-Ag) es muy similar, pero carece de los grupos de ácido siálico. Tal CDR3 de HC corta crea una ranura amplia en la que se podría unir un hidrato de carbono, y tales Abs son muy muy raros en bibliotecas de anticuerpos existentes. De este modo, las bibliotecas actuales no producen una gran variedad de ligantes potenciales a hidratos de carbono.

Ab 1B1 es el mAAb murino que se une a GBS-Ag; Ab 1QFU es el mAAb que tiene una estructura tridimensional conocida y la secuencia más cercana; y 1NSN es un anticuerpo de estructura tridimensional conocida que tiene una CDR3 de HC de longitud 4. El examen de una estructura de HC 3-23 da una distancia de C α de R₉₄ (que termina FR3) al C α de W₁₀₄ (que comienza FR4) de ~10 Å. La CDR3 de 1B1 (NWDY (SEQ ID NO:29)) muestra que los AAs no necesitan tener solamente grupos laterales pequeños o ser mayoritariamente de glicina. Tres aminoácidos (AAs) pueden puentear 10 Å, aunque PPP puede no funcionar. De hecho, se han obtenido unos pocos Fabs con CDR3s tan cortas como 3 AAs, pero son muy raros.

Aunque se han descrito CDR3s de HC cortas y muy cortas, nadie ha sugerido obtener una biblioteca de Ab que tenga muchos miembros (por ejemplo, mayor que alrededor de 50%, alrededor de 60%, alrededor de 70%, alrededor de 80%, alrededor de 90%, o alrededor de 95% de miembros) con CDR3s de HC cortas (por ejemplo, CDR3s de HC de 3 a 5 aminoácidos). Un enfoque para construir una biblioteca eficaz es diseñar en primer lugar secuencias de aminoácidos que podrían surgir del acoplamiento de V-J o V-D-J). Para una longitud de 3, 4 ó 5 de CDR3, se comienza con las secuencias de aminoácidos mostradas en la Tabla 7. Por ejemplo, la secuencia V-3JH1 muestra el extremo C-terminal de 3-23 FR3 (TAVYYCAK (SEQ ID NO:30)) seguido de JH1, que se ha recortado en el extremo N-terminal hasta tres aminoácidos antes de Trp-Gly que comienza FR4. V-3JH2 muestra el extremo de FR3 seguido del JH2 recortado. La secuencia que sigue V-3JH6 se construye uniendo FR4 a un trímero tomado de un segmento D humano, seguido de la región FR4 de un segmento JH humano. 3D3-3.3.2 sería un trímero del segmento D3-3, tercer marco de lectura partiendo en el segundo aminoácido. 5D5-12.2.3 es un pentámero de D5-12 en el marco de lectura 2 partiendo del aminoácido 3. Algunos de los segmentos D de línea germinal contienen codones de parada, aunque aparecen en anticuerpos naturales cuando los codones de parada se eliminan por edición: aquí se supone que el cambio más probable para los codones TAA y TAG es una Tyr (Y), y que las paradas TGA están muy probablemente mutadas a Trp (W). La Tabla 20 muestra las secuencias de aminoácidos de los segmentos D humanos; los tipos de codones de parada se indican mediante el uso de * para TAG, @ para TAA, y \$ para TGA. En la Tabla 11 hay 266 trímeros distintos que se pueden construir a partir de segmentos D humanos. Las paradas TAA y TAG se han cambiado a Tyr mostrado como "y" (es decir, minúsculas). Éstos también se podrían cambiar a Ser, Cys, Phe, Gln, Lys, o Glu mediante cambios de bases individuales. Todos los TAG se podrían cambiar mediante cambios de bases individuales a Trp, así como también Tyr, Gln, Lys, Glu, Ser, y Leu. La Tabla 12 muestra los 266 tetrámeros distintos que se pueden obtener recortando segmentos D humanos. La Tabla 13 muestra los 215 pentámeros que se pueden obtener recortando segmentos D humanos. La Tabla 14 muestra los 155 hexámeros que se pueden obtener recortando segmentos D humanos. Las bibliotecas a construir tienen diversidad sustancial en CDR1 de HC y CDR2 de HC. La diversidad de secuencia de CDR3 de HC puede ser menos importante que tener una secuencia corta pero aceptable. La diversidad de segmentos JH o fragmentos (por ejemplo, 3 o más aminoácidos) de segmentos D proporciona secuencias que se podrían construir mediante el sistema inmunitario humano y de este modo es menos probable que sean inmunogénicas.

En una realización, los trímeros, tetrámeros y pentámeros que contienen una Cys se eliminan.

En una realización, los trímeros, tetrámeros y pentámeros que contienen una Cys o proceden de un fragmento D que contiene una parada se eliminan.

Las bibliotecas cortas construidas usando los trímeros de la Tabla 11, tetrámeros de la Tabla 12, pentámeros de la Tabla 13 tienen diversidad sustancial: 266, 266, y 215 respectivamente. Esto se ha de comparar con el número de péptidos de estas longitudes: 8000, 160000, y 3200000 respectivamente.

V-3D1-1.1.1-JH1 contiene la porción final de FR3 seguido de tres aminoácidos de D1-1 (RF1), es decir, GTT (SEQ ID NO:257). V-3D1-1.2-JH1 usa los aminoácidos 2-4 de D1-1 (RF1) como la CDR3 progenitora. V-3D3-3.3.3-JH2

muestra el extremo de FR3 seguido de los aminoácidos 3-5 de D3-3 (RF 3). La invención comprende cualquier secuencia de aminoácidos que comprende FR3:: (tres, cuatro, o cinco AAs libres de parada de un segmento D humano)::FR4 de un JH humano. Los fragmentos de regiones D que contienen restos de Cys no emparejados se prefieren menos que aquellos que están libres de restos de Cys no emparejados. En V-5JH3, hay una Tyr mostrada como "y" debido a que JH3 tiene sólo 4 codones antes de los codones para Trp-Gly que definen el comienzo de FR4. V-5JH4 tiene un Ser mostrado como "s" por la misma razón. Si se usa titubeo, el nivel preferido de pureza está entre 0,75 y 0,90. La invención comprende las secuencias V-3JH1 a V-3JH6, V-4JH1 a V-4JH6, y V-5JH1 a V-5JH6, y las bibliotecas que las contienen. La invención también comprende las secuencias en las que la región de CDR está sustituida por un segmento de 3, 4 ó 5 aminoácidos de una región D humana, y bibliotecas que las contienen. La invención comprende además ADN en el que la secuencia progenitora se ha mutado en la región de CDR3, y bibliotecas que lo contienen. Una realización preferida es aquella en la que el número promedio de cambios de bases por CDR3 es uno, dos, o tres. Los métodos de mutagénesis incluyen PCR con tendencia a error, titubeo, y titubeo digital.

Tabla 7: Secuencias de aminoácidos de CDR3s progenitoras de longitudes 3, 4, 5

Longitud 3	...FR3-----	CDR3-	FR4-----	
V-3JH1	TAVYYCAK	FQH	WGQGTTLVTVSS	(SEQ ID NO:31)
V-3JH2	TAVYYCAK	FDL	WGRGTLVTVSS	(SEQ ID NO:32)
V-3JH3	TAVYYCAK	FDI	WGQGTMTVTVSS	(SEQ ID NO:33)
V-3JH4	TAVYYCAK	FDY	WGQGTTLVTVSS	(SEQ ID NO:34)
V-3JH5	TAVYYCAK	FDP	WGQGTTLVTVSS	(SEQ ID NO:35)
V-3JH6	TAVYYCAK	MDV	WGQGTTLVTVSS	(SEQ ID NO:36)
V-3D1-1.1.1-JH1	TAVYYCAK	GTT	WGQGTTLVTVSS	(SEQ ID NO:37)
V-3D1-1.1.2-JH1	TAVYYCAK	TTG	WGQGTTLVTVSS	(SEQ ID NO:38)
V-3D3-3.3.3-JH2	TAVYYCAK	IFG	WGRGTLVTVSS	(SEQ ID NO:39)

Longitud 4				
V-4JH1	TAVYYCAK	YFQH	WGQGTTLVTVSS	(SEQ ID NO:40)
V-4JH2	TAVYYCAK	YFDL	WGRGTLVTVSS	(SEQ ID NO:41)
V-4JH3	TAVYYCAK	AFDI	WGQGTMTVTVSS	(SEQ ID NO:42)
V-4JH4	TAVYYCAK	YFDY	WGQGTTLVTVSS	(SEQ ID NO:43)
V-4JH5	TAVYYCAK	WFDP	WGQGTTLVTVSS	(SEQ ID NO:44)
V-4JH6	TAVYYCAK	GMDV	WGQGTTLVTVSS	(SEQ ID NO:45)
V-4D3-10.1a-JH2	TAVYYCAK	LLWF	WGRGTLVTVSS	(SEQ ID NO:46)

Longitud 5				
V-5JH1	TAVYYCAK	EYFQH	WGQGTTLVTVSS	(SEQ ID NO:47)
V-5JH2	TAVYYCAK	WYFDL	WGRGTLVTVSS	(SEQ ID NO:48)
V-5JH3	TAVYYCAK	yAFDI	WGQGTMTVTVSS	(SEQ ID NO:49)
V-5JH4	TAVYYCAK	sYFDY	WGQGTTLVTVSS	(SEQ ID NO:50)
V-5JH5	TAVYYCAK	NWFDP	WGQGTTLVTVSS	(SEQ ID NO:51)
V-5JH6	TAVYYCAK	YGMDV	WGQGTTLVTVSS	(SEQ ID NO:52)

V-5D2-8.2a-JH2

TAVYYCAK

DIVLM

WGRGTLVTVSS

(SEQ ID NO:53)

Tabla 8: ADN que codifica V-5D2-8.2a-JH2 para titubeo

```

!                                     CDR3.....
!   A   E   D   T   A   V   Y   Y   C   A   K   D   I   V   L   M
! gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aag jez ezq jzz qzz ezj
!
!   W   G   Q   G   T   T   V   T   V   S   S   (SEQ ID NO:54)
!   tgg ggc cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:55)
!
!   BstEII...

```

Como alternativa, se podrían sintetizar tres segmentos de ADN que corresponden a la región de *XbaI* a *BstEII* y que tienen el resto 94 que es K o R seguido de 3, 4, ó 5 codones NNK, seguido de WG... de FR4. La variación permitida es $20^3 + 20^4 + 20^5 = 3.368.000$. Tras la amplificación, estas moléculas de ADN se mezclarían en la relación 1:10:100 (de manera que las secuencias más cortas están relativamente sobremuestreadas) y se clonarían en el fagómido que codifica la biblioteca con diversidad de CDR1/2 de HC. Una biblioteca de 1×10^9 daría diversidad significativa y permitiría el aislamiento de anticuerpos que se unen a dianas que tienen protuberancias pequeñas a medias. Por ejemplo, diversos hidratos de carbono, bucles de proteínas que no están bien ordenados (tales como GPCR's) se pueden beneficiar de una ranura en el anticuerpo creado al tener una CDR3 de HC muy corta. También se puede construir una biblioteca lambda. La relación de secuencias de AA es 1:20:400, y puede ser importante para muestrear las secuencias más cortas de forma más densa. La obtención de una ondonada amplia, grande, en el Ab puede requerir exactamente una CDR3 de 3 AAs, pero con una CDR3 de 4 AAs probablemente se tiene más flexibilidad, y con 5 AAs tiene incluso más flexibilidad. En este Ejemplo, se usa la versión JH6 de FR4 procedente del motivo WG en adelante.

Se puede seleccionar de nuestra biblioteca kappa actual una colección de, por ejemplo, 25 cadenas ligeras kappa que son a) línea germinal en las regiones del marco, b) muestran diversidad adecuada en las CDRs, y c) son de los tipos que producen bien y se emparejan bien con 3-23. Estas LCs se obtendrán en *E. coli* de un vector que posee *Kan^R* y ninguna señal de empaquetamiento del fago. Se podría construir entonces nuestra biblioteca de HC en un vector fágico que no tiene ninguna LC. HC y LC se cruzarán infectando las células productoras de LC con el fago de HC. El fago de HC que se selecciona se puede combinar con la LC de la célula que produce fago ELISA⁺, o las HC's se pueden clonar en pMID21 que tienen toda la diversidad de LC. Como alternativa, la HC seleccionada se puede mover en pHCSK85 y se puede usar con ROLIC para combinarse con todas las LCs de nuestra colección. También se podrían LCs lambda. De este modo, una biblioteca de 1×10^9 HC en fago se puede expandir en una biblioteca Fab de $1,2 \times 10^{11}$ ($1 \times 10^9 \times 117$). Si se combinan 1×10^7 CDR1-2s con 10^6 CDR3s de HC, se podría obtener una biblioteca de 5×10^7 en la que cada CDR3 se acopla con 50 CDR1-2s. Una biblioteca de 5×10^7 HC's en fago podría dar resultados similares a una biblioteca del estilo antiguo de 6×10^9 .

Tabla 1: Diseños de CDR3s de HC ejemplares muy cortas

```

C3XXX
!   ADN scab      S   R   D   N   S   K   N   T   L   Y   L   Q   M   N   S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
!   XbaI...
!
!                                     CDR3.....
!   L   R   A   E   D   T   A   V   Y   Y   C   A   K|R cualquiera cualquiera cualquiera W   G
! TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk tgg ggc-
!
!   Q   G   T   T   V   T   V   S   S   (SEQ ID NO:56)
!   cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:57)
!   BstEII...
!
! (C3XXX) 5'-T|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk tgg ggc cag ggt act ac-3'
! (SEQ ID NO:58)
! (ON_5) 5'-AcTggAgAcggTgAccgTAGTAcccTggcccca-3' ! 33 bases (SEQ ID NO:58)
! (ON_5 es complemento inverso de 5'-tgg ggc cag ggt act acG GTC ACC gtc tcc
! agt-3' (SEQ ID NO:59))
! Úsense ON-1 y ON-3 como se muestra más abajo
! -----
!

```

```

C3X4
! ADN scab      S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
!              XbaI...
!
!              CDR3.....
! L R A E D T A V Y Y C A K|R cualquiera cualquiera cualquiera cualquiera W
|TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk tgg-
!
! G Q G T T V T V S S (SEQ ID NO:60)
ggc cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:61)
!              BstEII...
!
(C3X4)5'-GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk tgg-
ggc cag ggt act ac-3' (SEQ ID NO:62)
! Úsense ON-1, ON-3 y ON-5
!-----
C3X5
! ADN scab      S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agc-
!              XbaI...
!
!              CDR3.....
! L R A E D T A V Y Y C A K|R cualquiera cualquiera cualquiera cualquiera cualquiera
|TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk nnk-
!
! W G Q G T T V T V S S (SEQ ID NO:63)
tgg ggc cag ggt act acG GTC ACC gtc tcc agt-3' (SEQ ID NO:64)
!              BstEII...
!
(C3X5)5'-GCA|GtT|taT|taC|tgc|gct aRg nnk nnk nnk nnk nnk tgg-
ggc cag ggt act ac-3' (SEQ ID NO:65)
!-----
aRg codifica K o R

```

Como alternativa, la diversidad de HC actual se puede clonar en DY3F87HC, y la diversidad de CDR3 descrita anteriormente se clona en esa diversidad como fragmentos XbaI – BstEII. Una biblioteca de, por ejemplo, 25 LC se clona en pLCSK23 y se usa para crear una estirpe celular en *E. coli* TG1. Estas células se infectan con el fago DY3F87HC que posee la nueva diversidad de CDR3 (y CDR1-2) de HC. El fago obtenido de esta infección se selecciona para la unión a una diana deseada. Después de dos a cuatro rondas de selección, las HCs seleccionadas se transfieren a pHCSK22 y se usan para crear una estirpe celular que se puede usar con ROLIC para combinar la HC seleccionada con todas las LCs en la biblioteca de LCs de ROLIC. De este modo, una biblioteca de 1. E 9 puede dar Abs que normalmente requerirían la construcción de una biblioteca de 1. E 16 (suponiendo una diversidad de LC de 1. E 7).

Ejemplo profético 2: Bibliotecas con CDR3s de HC muy largas

Sidhu et al. (J Mol Biol. 2004 338:299-310. y solicitud US 20050119455A1) dan a conocer Abs de alta afinidad seleccionados de una biblioteca en la que sólo se permiten Y y S en las CDRs que se limitaron en longitud a 20 aminoácidos. Puede ser posible generar Abs de alta afinidad a partir de una biblioteca que tiene CDR3s de HC con una o más de las siguientes formas de diversidad: a) varios (pero no todos) sitios que permiten Y o S, b) que incluyen 4-6 codones NNK, c) que introducen segmentos D (con o sin diversificación en el D), y/o d) usan PCR con tendencia a error. Se ha muestreado ya el espacio de Ab en el que la CDR3 de HC está en el intervalo -8 a ~22, con una longitud de la mediana de 13. De este modo, son posibles bibliotecas en las que CDR3 de HC tiene ~23 AAs o ~35 AAs, y pueden ser ventajosas con ciertos tipos de dianas. Por ejemplo, GPCRs son proteínas de la membrana integrales con siete segmentos helicoidales que atraviesan la bicapa lipídica de la célula que se piensa que tienen múltiples estados. Un anticuerpo que tiene una CDR3 de HC muy larga podría formar una protuberancia que se ajusta en el canal formado por las siete hebras. El hallar Abs que se unen a GPCRs ha sido difícil, y la construcción intencionada de bibliotecas en las que todos los miembros tienen CDR3s de HC muy largas puede mejorar este problema. Las longitudes se pueden hacer en cierto modo variables, digamos 23, 24, ó 25 en una biblioteca y 33, 34 ó 35 en una segunda.

Más abajo se da un número de diseños representativos. La CDR3 se ha roto y se ha generado diversidad que permite que las diversas partes tengan relaciones que difieren entre sí dependiendo del valor de X. Se ha usado un JH1 de longitud completa, y en algunos diseños la diversidad permitió diversidad en la parte CDR3 de JH1. Se podrían usar otros JHs. En los diseños, los segmentos D son ricos en Y o tienen un bucle de disulfuro rico en S. Las secuencias de aminoácidos de los segmentos D humanos se muestran en la Tabla 3. Se han variado en particular los lugares en los que la región D tiene S o Y o se permiten otras combinaciones. La Tabla 4 muestra las secuencias de aminoácidos de regiones J humanas.

Cada una de las bibliotecas se podría construir en al menos cuatro formas: 1) en primer lugar se sintetiza ADN que codifica una secuencia de aminoácidos particular y se somete a PCR con tendencia a error, 2) la biblioteca se puede sintetizar mediante titubeo o con mezclas de nucleótidos, 3) la biblioteca se puede construir usando titubeo digital, y 4) las rutas (2) ó (3) se podrían seguir mediante PCR con tendencia a error. Como un ejemplo de la ruta (1), en el

Se podría construir una biblioteca de HCs con CDR3 con longitud 23 de, por ejemplo, 2×10^9 miembros y una segunda biblioteca con CDR3s de HC de longitud ~ 35 que también tiene 2×10^9 miembros. Como alternativa, el ADN se podría mezclar para construir una biblioteca de 4×10^9 .

Tabla 4: Secuencias de aminoácidos de JH humano

	H3	
	CDR3	
	100	110
	I	I
JH1	---AEYFQHWGQGTTLVTVSS	(SEQ ID NO:66)
JH2	---YWYFDLWGRGTLVTVSS	(SEQ ID NO:67)
JH3	-----AFDIWGGGTMVTVSS	(SEQ ID NO:2)
JH4	-----YFDYWGGGTLVTVSS	(SEQ ID NO:1)
JH5	---NWFDPPWGGGTLVTVSS	(SEQ ID NO:68)
JH6	YYYYYGMDVWGGGTTVTVSS	(SEQ ID NO:3)

En cada uno de los siguientes diseños, la secuencia de aminoácidos comienza con YYCA(K/R), que es el extremo de FR3. FR4 comienza con WG y se muestra en negrita.

Diseño 1

SEQ ID NO:898 comprende el extremo de FR3 unido a dos restos (DY) de los tipos encontrados a menudo en la secuencia de relleno que el sistema inmunitario coloca entre V y D. A esto le sigue D2-2.2, preferido debido a que tiene un bucle de disulfuro y es rico en restos de Ser y Tyr. A esto le sigue YGYSY, que es rica en restos de Tyr y Ser, que es seguido de JH1 de longitud completa.

```

XX::D2-2.2::XX::JH1
      1 1 2 2
FR3 1 5 0 5 0 3FR4
YYCAK DYGYCSSTSCYTYGYSYAEYFQHWGQGTTLVTVSS (SEQ ID NO:898)
YYCAK XXGYCSXXSCYTXYGYSYAEYFQHWGQGTTLVTVSS (SEQ ID NO:69)
      R GYCSSTSCYT AEYFQHWGQGTTLVTVSS (JH1)
      (SEQ ID NO:70) (SEQ ID NO:66)
      1 1 1 1
9 9 0 0 0 1
4 5 0 2abcdefghijklmnop3 0

```

Diversidad de aminoácido = 1,28 E 8
 Diversidad de ADN = 2,15 E 9
 Libre de parada = 83%
 Libre de Cys innecesaria = 83%
 Libre de parada y Cys = 68%

El Diseño 1 (C23D222) tiene 94 que son R o K, después 2 Xs, D2-2 en un segundo marco de lectura con dos Xs en el bucle, seguido de dos Xs, y JH1. El 2º marco de lectura de D2-2 tiene un bucle cerrado con disulfuro en el que se ha introducido diversidad en dos puntos. Esta CDR3 tiene una longitud de 23. Usando los cebadores que incluyen ADN hasta ...YYCA y de WGQG..., se podría llevar a cabo la PCR con tendencia a error sobre la CDR3 antes de amplificar a XbaI y BstEII para la clonación en la biblioteca de LC kappa y CDR1/2 de HC. De este modo, se permitirá variar algunos de los AAs que se muestran como fijos. Los AAs que son parte de la región de solapamiento de PCR se reforzarán mediante la PCR sin tendencia a error final. La PCR con tendencia a error no es una parte necesaria del diseño.

C23D222JH1
 ! ADN scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
 XbaI...
 ! L R A E D T A V Y Y C A KIR
 |TTA|AGg|gct|gag|gaT|aCT|GCA|GtI|taT|taC|tgc|gct aRg -
 ! CDR3-----
 ! X X D2-2 RF2..... X X JH1..
 ! cualquiera cualquiera G Y C S cualquiera cualquiera S C Y T cualquiera cualquiera Y S Y A
 ! nnk nnk ggt tat tgt tcc nnk nnk tct tgc tat act nnk nnk tat tcc tac gct-
 ! CDR3-----
 ! E Y F Q H
 ! gaa tat ttc cag cac-
 ! W G Q G T L V T V S S (SEQ ID NO : 71)
 ! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO : 72)
 ! BstEII...
 (ON_C23D222) 5'-GCA|GtI|taT|taC|tgc|gct aRg nnk nnk ggt tat tgt tcc nnk-
 nnk tct tgc tat act nnk nnk tat tcc tac gct gaa tat ttc cag cac-
 tgg ggc cag ggt act ct-3' ! 107 bases (SEQ ID NO : 73)
 (ON_1) 5'-GCA|GtI|taT|taC|tgc|gct-3' (SEQ ID NO : 74)
 (ON_2) 5'-AgAgTAcccTggccccAgAcgTccATAccgTAATAgT-3' ! 37 bases (SEQ ID NO : 75)
 (ON_2 es complemento inverso de 5'-ac tat tac ggt atg gac gtc tgg
 ggc cag ggt act ct-3') (SEQ ID NO : 76)
 (ON_3) 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|
 aac|agC|TTA|AGg|gct|gag|gaT|aCT|GCA|GtI|taT|taC|tgc|gct-3' (SEQ ID
 NO : 77)
 (ON_4) 5'-AcTggAgAcggTgAccAgAgTAcccTggccccA-3' ! 33 bases (SEQ ID NO : 78)
 (5'-tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' [RC] (SEQ ID NO : 79))

Diseño 2

1 5 0 5 2 2
 1 5 0 5 0 3
 YYCAK GSYYYGSGSYYNVDSSYYAEYFQHWGQGLVTVSS (SEQ ID NO:899)
 YYCAK XYYYYGXGSXYNXXSYAEYFQHWGQGLVTVSS (SEQ ID NO:80)
 R YYYGSGSYYN AEYFQHWGQGLVTVSS (JH1)
 (SEQ ID NO:81) (SEQ ID NO:66)

Diversidad de aminoácido = 1,28 E 8
 Diversidad de ADN = 2,15 E 9
 Libre de parada = 83%
 Libre de Cys innecesaria = 83%
 Libre de parada y Cys = 68%

- 5 El Diseño 2 (C23D310) tiene 94 como R o K, dos Xs, D3-10 (RF2) con restos 5° y 8° cambiados a X, 2 Xs, SY Y, y JH1. La CDR3 tiene una longitud de 23 AAs, y se podría diversificar adicionalmente mediante el uso de PCR con tendencia a error.

C23D310JH1
 ! ADN scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
 XbaI...
 ! L R A E D T A V Y Y C A KIR
 |TTA|AGg|gct|gag|gaT|aCT|GCA|GtI|taT|taC|tgc|gct aRg -
 ! CDR3-----
 ! cualquiera cualquiera Y Y Y G cualquiera G S cualquiera Y N cualquiera cualquiera S Y Y
 ! nnk nnk tac tac tat ggt nnk ggc tct nnk tac aat nnk nnk tct tat tac
 ! A E Y F Q H
 ! gct gag tac ttt caa cat
 ! JH1.....
 ! W G Q G T L V T V S S (SEQ ID NO:82)
 ! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:83)
 ! BstEII...

(C23D310) 5'-GCAIGtTItaTItaCItgcIqct aRg nnk nnk tac tac tat ggt nnk ggc-
tct nnk tac aat nnk nnk tct tat tac gct gag tac ttt caa cat tgg ggc cag-
ggt act ct-3' (SEQ ID NO:84)
ON_1, ON_2, ON_3, y ON_4 como antes.

Diseño 3

		1	1	2	2
1	5	0	5	0	3

YYCAK DYYYYGSGSYNSDSYSAEYFQHWGQGLVTVSS (SEQ ID NO:900)
YYCAK XZYZZGZGXYNZXZYZAXZFQHWGQGLVTVSS (SEQ ID NO:84)
R YYYGSGSYN AEYFQHWGQGLVTVSS (JH1)
(SEQ ID NO:81) (SEQ ID NO:66)

Diversidad de aminoácido = 1,64 E 8
Diversidad de ADN = 1,07 E 9
Libre de parada = 88%
Libre de Cys innecesaria = 88%
Libre de parada y Cys = 77%

- 5 Diseño 3 (C23D310B) tiene 94 como R o K, XZ, D3-10 (RF2) con el resto 2°, 3°, 5°, y 7° como Z(Y|S) y 8° cambiado a X, ZXZYZ, y JH1 (con el E cambiado a X). Z es Y o S. La CDR3 tiene una longitud de 23 AAs, y se podría diversificar adicionalmente mediante el uso de PCR con tendencia a error.

A V Y Y C A R|K cualquier Y|S Y Y|S Y|S G Y|S
(C23D310b) 5'-GCAIGtTItaTItaCItgcIqct aRg nnk tmc tac tmc tmt ggt tmc ggc-
Y|S cualquier Y N Y|S cualquier Y|S Y Y|S A cualquier Y|S F Q H W G Q
tmt nnk tac aat tmt nnk tmc tat tmc gct nnk tmc ttt caa cat tgg ggc cag-
G T L (SEQ ID NO:85)
ggt act ct-3' (SEQ ID NO:86)

ON_1, ON_2, ON_3, y ON_4 como antes.

Diseño 4

		1	1	2	2	2	3	3
1	5	0	5	0	3	5	0	5

YYCAK YYSFSYYPYYYDSSGYYYGYSDYSYSAEYFQHWGQGLVTVSS (SEQ ID NO:901)
YYCAK YYSXSYYXZYDSZGYZYXYYSXZYZZAZZFQHWGQGLVTVSS (SEQ ID NO:87)
R YYYDSSGYYY AEYFQHWGQGLVTVSS (JH1)
(SEQ ID NO:88) (SEQ ID NO:66)

1	1	1	1
9	9	0	0
4	5	0	2

abcde fghij klmnopq rstuvwxyab3

Diversidad de aminoácido = 1,64 E 8
Diversidad de ADN = 1,07 E 9
Libre de parada = 88%
Libre de Cys innecesaria = 88%
Libre de parada y Cys = 77%

10

El Diseño 4 tiene CDR3 de longitud 35. El resto 94 puede ser K o R, después YYS::X::SYY::X::D3-22 (2° RF con un S como X y 3 Zs)::X::YYS::X::YZZZ::JH1 (con 2 Zs). Para añadir más diversidad, se podría usar PCR con tendencia a error.

C35D322JH1
 ! ADN scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
 ! XbaI...
 !
 ! L R A E D T A V Y Y C A KIR
 ! ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg -
 !
 ! CDR3-----
 ! Y Y S cualquiera S Y Y cualquiera Y Y|S Y D S Y|S G Y Y|S Y
 ! tac tat tcc nnk tct tac tat nnk tat tmt tac gat agt tmt ggt tac tmc tat
 !
 ! cualquiera Y Y S cualquiera Y Y|S Y Y|S Y|S Y|S A Y|S Y|S F Q H
 ! nnk tac tat agc nnk tat tmc tac tmc tmt tmc gct tmt tmc ttc caa cac
 !
 ! W G Q G T L V T V S S (SEQ ID NO:89)
 ! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:90)
 ! BstEII...
 !
 ! (c35d322b) 5'-GCA|GtT|taT|taC|tgc|gct aRg tac tat tcc nnk tct tac tat nnk-
 ! tat tmt tac gat agt tmt ggt tac tmc tat nnk tac tat agc nnk tat tmc tac-
 ! tmc tmt tmc gct tmt tmc ttc caa cac tgg ggc cag ggt act ct-3' (SEQ ID NO:91)
 ! ON_1, ON_2, ON_3, y ON_4 como antes.

Diseño 5

1 5 0 5 0 3
 YYCAK SSGYCSSTSCYTG VY Y Y A E Y F Q H W G G C T L V T V S S (SEQ ID NO:902)
 YYCAK ZZGZCZZXZCZTXXYZYXZYFQHWG G C T L V T V S S (SEQ ID NO:92)
 R GYCSSTSCYT AEYFQHWG G C T L V T V S S (JH1)
 (SEQ ID NO:70) (SEQ ID NO:66)
 Diversidad de aminoácido = 1,64 E 8
 Diversidad de ADN = 1,07 E 9
 Libre de parada = 88%
 Libre de Cys innecesaria = 88%
 Libre de parada y Cys = 77%

5 El Diseño 5 (C23D222b) es como el diseño 1, pero usa muchos codones variables Z (Y o S). Esta CDR3 tiene una longitud de 23.

C23D222JH1b
 ! ADN scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
 ! XbaI...
 !
 ! L R A E D T A V Y Y C A KIR
 ! ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg -
 !
 ! CDR3-----
 ! Y|S Y|S G Y|S C Y|S Y|S cualquiera Y|S C Y|S T cualquiera cualquiera Y Y|S Y cualquiera
 ! tmc tmt ggt tmt tgc tmc tmt nnk tmt tgt tmc acc nnk nnk tat tmt tac nnk
 !
 ! Y|S Y F Q H
 ! tmt tat ttc cag cac
 !
 ! W G Q G T L V T V S S (SEQ ID NO:93)
 ! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:94)
 ! BstEII...
 !
 ! (C23D222JH1b) 5'-GCA|GtT|taT|taC|tgc|gct aRg tmc tmt ggt tmt tgc tmc tmt-
 ! nnk tmt tgt tmc acc nnk nnk tat tmt tac nnk tmt tat ttc cag cac tgg ggc-
 ! cag ggt act ct-3' (SEQ ID NO:95)

Diseño 6

```

      1      1      2 2 2      3      3
      1 5      0 5      0 3 5      0 5
YYCAK SYDYGYCSTSCYTYYSVSYSSYSSYAEYFQHWGQGLVTVSS (SEQ ID NO:903)
YYCAK ZYXZYGZCZXXSCZTYZSZXZYSZYSSZYAEZFOHWGQGLVTVSS (SEQ ID NO:96)
      R      GYCSSTSCYT D2-2.2      AEYFQHWGQGLVTVSS (JH1)
              (SEQ ID NO:70)              (SEQ ID NO:66)

```

Diversidad de aminoácido = 2,00 E 8
 Diversidad de ADN = 5,37 E 8
 Libre de parada = 91%
 Libre de Cys innecesaria = 91%
 Libre de parada y Cys = 83%

```

C35D222JH1
!
! ADN scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
! XbaI...
!
! L R A E D T A V Y Y C A KIR
! ITTA|AGg|gct|gag|gaT|aCT|GCA|Gt|T|taT|taC|tgc|gct aRg -
!
! CDR3-----
! Y|S Y cualquiera Y|S Y G Y|S C Y|S Y|S cualquiera S C Y|S T Y Y|S S
! tmt tac nnk tmc tac gcc tmt tgc tmt tmc nnk tCt tgt tmc acc tat tmt tcc
!
! Y|S cualquiera Y|S Y S cualquiera Y Y|S S Y|S Y A E Y F Q H
! tmt nnk tmc tat tct nnk tac tmc agt tmt tat gct gag tat ttc cag cac
!
! W G Q G T L V T V S S (SEQ ID NO:97)
! tgg gcc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:98)
! BstEII...

(C35D222JH1)5'-GCA|Gt|T|taT|taC|tgc|gct aRg tmt tac nnk tmc tac gcc tat- tgc tmt tmc
nnk tmt tgt tmc acc tat tmt tcc tmt nnk tmc tat tct nnk tac-
tmc agt tmt tat gct gag tat ttc cag cac tgg gcc cag ggt act ct-3' (SEQ ID NO:99)

```

Diseño 7

```

      1      1      2 2 2      3      3
      1 5      0 5      0 3 5      0 5
YYCAK YYSYGYCSTSCYTYSSSVSYSSYSSYAEYFQHWGQGLVTVSS (SEQ ID NO:904)
YYCAK ZYZZYGZCZXXCZTYZSZXZYSZYSSZYAEZJQBHWGQGLVTVSS (SEQ ID NO:100)
      R      GYCSSTSCYT D2-2.2      AEYFQHWGQGLVTVSS (JH1)
              (SEQ ID NO:70)              (SEQ ID NO:66)
(J=FSY, B=YHND, Ψ=EKQ)

```

Diversidad de aminoácido = 9,44 E 8
 Diversidad de ADN = 2,42 E 9
 Libre de parada = 93%
 Libre de Cys innecesaria = 93%
 Libre de parada y Cys = 88%

```

C35D222JH1B
!
! ADN scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
! XbaI...
!
! L R A E D T A V Y Y C A KIR
! ITTA|AGg|gct|gag|gaT|aCT|GCA|Gt|T|taT|taC|tgc|gct aRg -
!
! CDR3-----
! Y|S Y Y|S Y|S Y G Y|S C Y|S Y|S cualquiera Y|S C Y|S T Y Y|S S
! tmt tac tmc tmc tac gcc tmt tgc tmt tmc nnk tmt tgt tmc acc tat tmt tcc
!
! Y|S cualquiera Y|S Y S Y|S Y Y|S S Y|S Y A E|K Y|S F|S Q H|Y
! tmt nnk tmc tat tct tmt tac tmc agt tmt tat gct Vag tmt tHc cag Nac
!
! W G Q G T L V T V S S (SEQ ID NO:101)
! tgg gcc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:102)
! BstEII...

```

```

      1      1      2 2 2      3      3
      1 5 0 5 0 3 5 . 0 5
YYCAK SRSYYDYVWGSYRYTSSYSYSSYSSYAEYFQHWGQGLTVTVSS (SEQ ID NO:905)
YYCAK ZXZYZBZVWGZZRZTZSZXZYZZYZSZAψZFQHWGQGLTVTVSS (SEQ ID NO:103)
      R      YYDYVWGSYRYT D3-16.2      AEYFQHWGQGLTVTVSS (JH1)
              (SEQ ID NO:104)              (SEQ ID NO:66)
(J=FSY, B=YHND, ψ=EKQ)

```

Diversidad de aminoácido = 9,44 E 8
 Diversidad de ADN = 1,61 E 9
 Libre de parada = 93%
 Libre de Cys innecesaria = 93%
 Libre de parada y Cys = 88%

C34D316JH1A

```

!
! ADN scab S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
! XbaI...
!
! L R A E D T A V Y Y C A KIR
! TTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgclgct aRg -
!
! CDR3-----
! NID
! Y|S cualquiera Y|S Y Y|S Y|H Y|S V W G Y|S Y|S R Y|S T Y|S
! tmt nnk tmc tac tmt Nat tmt gtt tgg ggt tmt tmc cgt tmt act tmt
!
! S Y|S cualquiera Y|S Y Y|S Y|S Y|S Y Y|S S Y|S
! agt tmc nnk tmt tac tmc tmt tmc tat tmc agt tmt
!
! Q
! A, E|K Y|S F Q H
! GCT vag tmc ttc cag cat
!
! W G Q G T L V T V S S (SEQ ID NO:105)
! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:106)
! BstEII...
!
! (C34D316JH1A) 5'-GCA|GtT|taT|taC|tgclgct aRg tmt nnk tmc tac tmt Nat tmt-
! gtt tgg ggt tmt tmc cgt tmt act tmt agt tmc nnk tmt tac tmc tmt tmc tat-
! tmc agt tmt GCT vag tmc ttc cag cat tgg ggc cag ggt act ct -3' (SEQ ID NO:107)

```

Diseño 9

- 5 El Diseño 9 es como el 8, excepto que el segmento D se mueve a la derecha.

```

      1      1      2 2 2      3      3
      1 5 0 5 0 3 5 0 5
YYCAK YGYSSDSYSSYYDYVWGSYRYTYSSYYAEYFQHWGQGLTVTVSS (SEQ ID NO:906)
YYCAK ZXZZZXZYZZZYBZVWGZZRZTYZSZAψZFQHWGQGLTVTVSS (SEQ ID NO:108)
      R      D3-16.2      YYDYVWGSYRYT      AEYFQHWGQGLTVTVSS (JH1)
              (SEQ ID NO:104)      (SEQ ID NO:66)
(J=FSY, B=YHND, ψ=EKQ)

```

Diversidad de aminoácido = 1,31 E 8
 Diversidad de ADN = 5,37 E 8
 Libre de parada = 91%
 Libre de Cys innecesaria = 91%
 Libre de parada y Cys = 83%

C34D316JH1B

```

!   ADN scab   S R D N S K N T L Y L Q M N S
5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
!           XbaI...
!   L R A E D T A V Y Y C A KIR
!   ITTA|AGg|gct|gag|gaT|aCT|GCA|GtT|taT|taC|tgc|gct aRg -
!   CDR3-----
!   Y|S cualquier Y|S Y|S Y|S cualquier Y|S Y Y|S Y|S Y|S
!   tmt nnk tmc tmt tmc nnk tmt tac tmc tmt tmc
!
!   NID
!   Y Y|S Y|H Y|S V W G Y|S Y|S R Y|S T
!   tac tmt Nat tmt gtt tgg ggt tmt tmc cgt tmt act
!
!   Y Y|S S Y|S Y
!   tat tmc agt tmt tac
!
!   Q
!   A E|K Y|S F Q H
!   GCT vag tmc ttc cag cat
!
!   W G Q G T L V T V S S (SEQ ID NO:109)
!   tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:110)
!           BstEII...

```

(C35D316JH1B)

```

5'-GCA|GtT|taT|taC|tgc|gct aRg tmt nnk tmc tmt tmc nnk tmt tac tmc tmt tmc
tac tmt Nat tmt gtt tgg ggt tmt tmc cgt tmt act tat tmc agt tmt tac GCT vag
tmc ttc cag cat tgg ggc cag ggt act ct-3' (SEQ ID NO:111)

```

Diseño 10

```

          1      1      2      2
        1 5      0 5      0 4
YYCAK GSSYYGSGSYNSDYSAEYFQHWGQGLTVTVSS (SEQ ID NO:907)
YYCAK XZZYZZGZGXNZXZYZAXZFOHWGQGLTVTVSS (SEQ ID NO:112)
      R   YYYGSGSYNN   AEYFQHWGQGLTVTVSS (JH1)
          (SEQ ID NO:81)   (SEQ ID NO:66)

```

- 5 El Diseño 10 (C24D310B) es como el Diseño 3, pero la CDR3 tiene una longitud de 24. El Diseño 10 tiene 94 como R o K, XZZ, D3-10 (RF2) con el resto 2°, 3°, 5°, y 7° como Z(YIS) y 8° cambiado a X, ZXZYZ, y JH1 (con el E cambiado a X). Z es Y o S. La CDR3 tiene una longitud de 24 AAs, y se podría diversificar adicionalmente mediante el uso de PCR con tendencia a error.

```

(C24D310b) 5'-GCA|GtT|taT|taC|tgc|gct aRg nnk tmc tmc tac tmc tmt ggt tmc-
ggc tmt nnk tac aat tmt nnk tmc tat tmc gct nnk tmc ttt caa cat tgg ggc-
cag ggt act ct-3' (SEQ ID NO:113)

```

ON_1, ON_2, ON_3, y ON_4 como antes.

Diseño 11

```

          1      1      2      2
        1 5      0 5      0 5
10 YYCAR SSRSGYCTNGVCYRSGSYWYFDLWGRGTLTVTVSS (SEQ ID NO:907)
YYCAR ZZXZGZC32GVC23ZXZZ4Z12LWGRGTLTVTVSS (SEQ ID NO:114)
      K   GYCTNGVCYT YWYFDLWGRGTLTVTVSS D2-8.2 JH2
          (SEQ ID NO:115)   (SEQ ID NO:67)
(1=FYS(THT), 2=YHND(NAT), 3=ITKR(ANA), 4=LSW(TBG))

```

```

(C24D282) 5'-GCA|GtT|taT|taC|tgc|gct aRg tmc tmt nnk tmt ggt tmc tgc ana-
nat ggt gtc tgc tmt ana tmc nnk tmt tmt tbg tmt tmt nat ctg tgg ggc-
cag ggt act ct-3' (SEQ ID NO:116)

```

```

(C24D282.1) 5'-GCA|GtT|taT|taC|tgc|gct aRg tmc tmt nnk tmc ggt tmc tgc ana-
nat ggc gtc tgc tmt ana tmc nnk tmt tmt tbg tmt tmt nat ctg tgg ggc-
cag ggt act ct-3' (SEQ ID NO:117)

```

```

(C24D282.1) 5'-GCA|GtT|taT|taC|tgc|gct aRg tmc tmt nnk tmc ggt tmc tgc ana-
nat ggc gtc tgc t-3' (SEQ ID NO:118)(necesita R, M, N, K)

```

```

(C24D282.2) 5'-Ag AgT Acc cTg gcc cCa cAg ATN ADA AKA cVA AKA AKA MNN gKA TNT AKA gCa
gAc gcc ATN TNT gCa gKA Acc g-3' (SEQ ID NO:119) ! 75 bases
(5'-e ggt tmc tgc ana-
nat ggc gtc tgc tmt ana tmc nnk tmt tmt tbg tmt tmt nat ctg tgg ggc-
cag ggt act ct-3' (RC) (SEQ ID NO:120)(necesita N, M, K, B, H))

```

Diseño 12

```

      1      1      2      2      3      3
      1      5      0      5      0      5
YYCAR SSSYSYGXCTNGVCYTYSYSYSYSYWFDLLWGRGTLVTVSS (SEQ ID NO:908)
YYCAR ZZZZZZGZC32GVCZ3ZZZZYZZYZZ4212LWGRGTLVTVSS (SEQ ID NO:121)
      K      GYCTNGVCYT      YWFDLLWGRGTLVTVSS D2-8.2 JH2
      (SEQ ID NO:115)      (SEQ ID NO:67)
(1=FYS, 2=YHND, 3=ITKR, 4=LSW, Z=YS)

```

(C33D282TP) 5'-GCAIGtTItaTItaCItgcIgcT-3' (SEQ ID NO:909)
(C33D282BP) 5'-ag agt acc ctg gcc cca-3' (SEQ ID NO:910)
(C33D282) 5'-GCAIGtTItaTItaCItgcIgcT aRg tmt tmc tmc tmt tmt tmc ggt-
tmt tgt ana nat ggc gtg tgc tmt ana tmc tmc tmc tmt tat tmt tmc tat tmt-
tac tmt tmc tbg tmc tht nat ctg tgg ggc cag ggt act ct-3' (SEQ ID NO:122)
(C33D282F) 5'-GCAIGtTItaTItaCItgcIgcT agg tct tcc tac tat tcc tac ggt-
tat tgt aca aat ggc gtg tgc tat aca tac tcc tac tct tat tat tcc tat tct-
tac tct tac tgg tac ttt gat ctg tgg ggc cag ggt act ct-3' (SEQ ID NO:911)

Diseño 13

El Diseño 13 coloca un segmento D de línea germinal en el centro de un mar de Zs, de manera que se pueden obtener dos trozos de ADN que solapan a lo largo de la región constante. La CDR3 de HC tiene una longitud de 34, y la diversidad es $2^{23} \sim 8 \times 10^6$.

1 1 2 2 3 3
1 5 0 5 0 5 0 5
YYCAR SSSYSYSSGYCTNGVCYTYSSYYSSYYWYFDLWGRGTLVTVSS (SEQ ID NO:912)
YYCAR ZZZZZZZZZZGYCTNGVCYTZZZZZZZZWZF2LWGRGTLVTVSS (SEQ ID NO:123)
K GYCTNGVCYT WYFDLWGRGTLVTVSS D2-8.2 JH2
. (SEQ ID NO:115) (SEQ ID NO:67)

(2=YHND)

(C34D282.2A) 5'-GCAIGTTItaTItaCItgcIgc aRg tmt tmc tmc tmt tmt tmc tmc tmt-
tmc tmc ggt tat tgt act aac ggc gtt tgc tat act-3' (SEQ ID NO:124)

(C34D282.2B) 5'-Ag AgT Acc cTg gcc cCA cAg gTn gAA AKA cCA AKA AKA gKA-
gKA gKA gKA AKA AKA AgT ATA gCA AAC gcc gTT AgT AcA ATA-3' (SEQ ID NO:125)! 86
bases

(5'- tat tgt act aac ggc gtt tgc tat act tmt tmt tmc tmc tmc tmc-
tmt tmt tmt tgg tmt ttc Nac ctg tgg ggc cag ggt act ct-3' (SEQ ID
NO:126) [RC])

Diseño 14

El Diseño 14 es igual a 9, excepto que el segmento D es mayoritariamente línea germinal.

```

      1      1      2 2 2      3      3
      1 5      0 5      0 3 5      0 5
YYCAK  YSYYSGSYYSDYVWGSYRYTSDYSYYAEYFQHWGGQGLTVTVSS (SEQ ID NO:913)
YYCAK  ZZZZZZZZZZDYVWGSYRZTZZZZZZZAEZFQHWGGQGLTVTVSS (SEQ ID NO:127)
      R   D3-16.2 YYDVWGSYRYT      AEYFQHWGGQGLTVTVSS (JH1)
              (SEQ ID NO:104)      (SEQ ID NO:66)

```

(C34D316.2A) 5'-GCAIGtT|taT|taC|tgc|gct aRg tmt tmc tmc tmt tmt tmc tmc tmt-
tmc tmc tmc gat tat gtc tgg ggt act tat cgt-3' (SEQ ID
NO:128)
(C34D316.2B) 5'-Ag AgT Acc cTg gcc cCA ATg cTg gAA AKA cTc Agc gKA gKA gKA-
gKA gKA gKA AKA AgT gKA Acg ATA AgT Acc ccA gAc ATA ATc-3' (SEQ ID
NO:129) ! 86 bases
(5'-gat tat gtc tgg ggt act tat cgt tmc act tmt tmc tmc tmc tmc-
tmc tmc gct gag tmt ttc cag cat tgg ggc cag ggt act ct-3'
(SEQ ID NO:130) [RC])

Diseño 15

El Diseño 15 permite cierta diversidad en el solapamiento, 5 movimientos de vaivén de dos vías. Sólo hay 32 secuencias de solapamiento, e incluso si hay desemparejamientos, no cambiarán la diversidad permitida.

```

      1   1   2 2 2   3   3
      1   5   0   5   0 3 5   0   5
YYCAK SYDYSSSYDYVWGSYRYTSYSGDSYAAEYFQHWGQGLVTVSS (SEQ ID NO:914)
YYCAK ZZZZZZZZDZVWGZZRZTZZZZZZAEZFQHWGQGLVTVSS (SEQ ID NO:131)
      YYDYVWGSYRYT      AEYFQHWGQGLVTVSS
      (SEQ ID NO:104)      (SEQ ID NO:66)

(C35D316.2A) 5'-GCAIGtItaItaCItgcIgcT aRg tmt tmc tmc tmt tmt tmc tmc tmt-
      tmc tmc tmc gac tmt gtc tgg ggt tmc tmc cgt tmc acc t-3' (SEQ ID NO:
132)
(C35D316.2B) 5'-Ag AgT Acc cTg gcc cCA ATg cTg gAA AKA cTc Agc gKA gKA-
      gKA gKA gKA gKA gKA gKA ggT gKA Acg gKA gKA Acc cCA gAc AKA gTc gKA g-3'
(SEQ ID NO:133)
(5'-c tmc gac tmt gtc tgg ggt tmc tmc cgt tmc acc tmt tmc tmc-
      tmc tmc tmc tmc gct gag tmt ttc cag cat tgg ggc cag ggt acc ct-3' (SEQ ID
NO:134) [RC])

```

Diseño 16

El Diseño 16 proporciona una CDR3 de 35. Hay 4 movimientos de vaivén de dos vías en el solapamiento, y de este modo 16 secuencias.

```

      1   1   2 2 2   3   3
      1   5   0   5   0 3 5   0   5
YYCAK SSSYSSSYSGYCSGGSCYSSYYSSYSAEYFQGWGQGLVTVSS (SEQ ID NO:915)
YYCAK ZZZZZZZZGZCZGGZCZSZZZZZZAEZFQHWGQGLVTVSS (SEQ ID NO:135)
      R      GYCSGGSCYS 2-25.2 AEYFQHWGQGLVTVSSJH1
      (SEQ ID NO:136)      (SEQ ID NO:66)

```

```

(C35D225.2A) 5'-GCAIGtItaItaCItgcIgcT aRg tmt tmt tmt tmt tmt tmt tmt tmt-
      tmc tmc ggc tmc tgt tmc ggt ggc tmc tgc tmc tcc t-3' (SEQ ID NO:137)
(C35D225.2B) 5'-Ag AgT Acc cTg gcc cCA ATg TTg gAA AKA TTc Agc gKA gKA-
      gKA gKA gKA gKA gKA gKA ggA gCA gKA gcc Acc gKA AcA gKA gcc gKA g-3'
(SEQ ID NO:138)! 96 bases

```

5

Si se añaden C34D225.2A y C34D225.2B a la mezcla, entonces obtenemos las CRD3s de longitudes 33, 34, y 35.

```

(C34D225.2A) 5'-GCAIGtItaItaCItgcIgcT aRg tmt tmt tmt tmt tmt tmt tmt tmt-
      tmc tmc ggc tmc tgt tmc ggt ggc tmc tgc tmc tcc t-3' (SEQ ID NO:139)
(C34D225.2B) 5'-Ag AgT Acc cTg gcc cCA ATg TTg gAA AKA TTc Agc gKA gKA-
      gKA gKA gKA gKA gKA gKA ggA gCA gKA gcc Acc gKA AcA gKA gcc gKA g-3'
(SEQ ID NO:140)! 93 bases

```

Diseño 17

```

      1   1   2 2 2   3   3
      1   5   0   5   0 3 5   0   5
YYCAK YSSSYSSYDYVWGSYRYTSSSYSSYSSYSAEYFQGWGQGLVTVSS (SEQ ID NO:916)
YYCAK ZZZZZZZZDZVWGZZRZTZZZZZZAEZFQHWGQGLVTVSS (SEQ ID NO:141)
      R      YYDYVWGSYRYT D3-16.2 AEYFQHWGQGLVTVSS (JH1)
      (SEQ ID NO:104)      (SEQ ID NO:66)

```

```

(C35D3162A) 5'- GCAIGtItaItaCItgcIgcT aRg tmt tmt tmt tmt tmt tmt tmc gac-
      tmc gtc tgg ggt tmt tmc cgt tmt acc t-3' (SEQ ID NO:142)
(C35D3162B) 5'-Ag AgT Acc cTg gcc cCA gTg cTg gAA gKA cTc Agc gKA gKA gKA-
      gKA gKA gKA gKA gKA gKA gKA gKA ggT AKA Acg gKA AKA Acc cCA gAc-
      gKA gTc g-3' (SEQ ID NO:143)

```

10

Diseño 18

```

      1   1   2 2 2   3   3
      1   5   0   5   0 3 5   0   5
YYCAK SSSYSSSYDYVWGSYRYTSSSYSSYSSYSAEYFQGWGQGLVTVSS (SEQ ID NO:917)
YYCAK ZZZZZZZZDZVWGZZRZTZZZZZZAEZFQHWGQGLVTVSS (SEQ ID NO:144)
      R      YYDYVWGSYRYT D3-16.2AEYFQHWGQGLVTVSS (JH1)
      (SEQ ID NO:104)      (SEQ ID NO:66)

```

```

(C35D3162C) 5'- GCAIGtItaItaCItgcIgcT aRg tmt tmt tmt tmt tmt tmt tmc-
      tmc tmc tmc gac tmc gtc tgg ggt tmc tmc cgt tmc acc t-3' (SEQ ID NO:145)
82 bases
(C35D3162B) 5'-Ag AgT Acc cTg gcc cCA gTg cTg gAA gKA cTc Agc gKA gKA-

```

gKA gKA gKA gKA gKA gKA gKA ggT gKA Acg gKA gKA Acc cca gAc gKA-
gTc g-3' (SEQ ID NO:146)

Diseño 19

		1	1	2	2	2	3	3
1	5	0	5	0	3	5	0	5

YYCAK YSGDSYSSYYYDSSGYYYSSSYSSSYAEEYFQGWGQGLTVTVSS (SEQ ID NO:918)
 YYCAK ZZZZZZZZZZDSSGZZZZZZZZZZZZZZAEZFQHWGQGLTVTVSS (SEQ ID NO:147)
 R YYYDSSGYYY AEYFQHWGQGLTVTVSS (JH1)
 (SEQ ID NO:88) (SEQ ID NO:66)

	1	1		1	1
9	9	0	0	0	1
4	5	0	2	abc	defghijklmnopqrstuvwxyab3
					0

Diversidad de aminoácido = 6,7 E 7
 Diversidad de ADN = 6,7 E 7
 Libre de parada = 100
 Libre de Cys innecesaria = 100
 Libre de parada y Cys = 100%

- 5 El Diseño 19 tiene una CDR3 de longitud 35. El resto 94 puede ser K o R, el ZZZZZZZZZ::D3-22 (2° RF con seis Ys como Z)::ZZZZZZZZZZ::JH1 (con 1 Z). Para añadir más diversidad, se podría usar PCR con tendencia a error.

C35D322AJH1
 ! ADN scab S R D N S K N T L Y L Q M N S
 5'-ttc|act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|aac|agC-
 ! XbaI...
 !
 ! L R A E D T A V Y Y C A KIR
 ! TTA|AGg|gct|gag|gaT|aCT|GCA|GtI|taT|taC|tgc|gct aRg -
 !
 ! CDR3-----
 !
 ! YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS D S S G YIS YIS YIS
 ! tmc tmt tmc tmc tmt tmc tmt tmc tmc tmc tmc gac agc tcc ggc tmc tmc tmt
 !
 ! YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS YIS A E YIS F Q H
 ! tmc tmt tmc tmc tmt tmc tmt tmc tmc tmc gct gaa tmc ttc caa cac
 !
 ! W G Q G T L V T V S S (SEQ ID NO:148)
 ! tgg ggc cag ggt act ctG GTC ACC gtc tcc agt-3' (SEQ ID NO:149)
 ! BstEII...

(C35D322AJH1_T) 5'-GCA|GtI|taT|taC|tgc|gct aRg tmc tmt tmc tmc tmt-
 tmc tmt tmc tmc tmc tmc gac agc tcc ggc tmc tmc t-3' (SEQ ID NO:150)
 (C35D322AJH1_B) 5'-cAg AgT Acc cTg gcc cca gTg TTg gAA gKA TTc Agc gKA-
 gKA gKA gKA AKA gKA AKA gKA gKA gKA AKA gKA gKA gcc gga gCT gTc-
 gKA gKA g-3' (SEQ ID NO:151)

ON_1, ON_2, ON_3, y ON_4 como antes.

Diseño 20

		1	1	2	2	2	3	3
1	5	0	5	0	3	5	0	5

YYCAK YSSYSS YYYDSSGYYYSSSYSSSYSS YYYAEYFQGWGQGLTVTVSS (SEQ ID NO:919)
 YYCAK ZZZZZZ(Z)ZZZZDSSGZZZZZZZZZZ(Z)ZZZAEZFQHWGQGLTVTVSS (SEQ ID NO:152)
 R YYYDSSGYYY AEYFQHWGQGLTVTVSS (JH1)
 (SEQ ID NO:88) (SEQ ID NO:66)

	1	1		1	1
9	9	0	0	0	1
4	5	0	3	abc	defghijklmnop q rstuvwxya4
					0

Diversidad de aminoácido = 6,7 E 7
 Diversidad de ADN = 6,7 E 7
 Libre de parada = 100
 Libre de Cys innecesaria = 100
 Libre de parada y Cys = 100%

El Diseño 20 tiene CDR3s de longitudes 33, 34, ó 35. El resto 94 puede ser K o R, el ZZZZZZ(Z)ZZ::D3-22 (2º RF con seis Ys como Z)::ZZZZZZ(Z)ZZ::JH1 (con 1 Z). La combinación mediante PCR (C35D322AJH1_T), (C34D322AJH1_T), (C35D322AJH1_B), y (C34D322AJH1_B) permite diversidad en longitud así como en secuencia.

```
(C35D322AJH1_T) 5'-GCAIGtTitaTitaCItgcIgcT aRg tmc tmt tmc tmc-
tmt tmc tmt tmc tmc tmc gac agc tcc ggc tmc tmc t-3' (SEQ ID NO:153)
(C34D322AJH1_T) 5'-GCAIGtTitaTitaCItgcIgcT aRg tmc tmc tmc tmt-
tmc tmt tmc tmc tmc tmc gac agc tcc ggc tmc tmc t-3' (SEQ ID NO:154)
(C35D322AJH1_B) 5'-cAg AgT Acc cTg gcc ccA gTg TTg gAA gKA TTc Agc gKA-
gKA gKA gKA AKA gKA AKA gKA gKA AKA gKA gKA ggc ggA gCT gTc-
gKA gKA g-3' (SEQ ID NO:920)
(C34D322AJH1_B) 5'-cAg AgT Acc cTg gcc ccA gTg TTg gAA gKA TTc Agc gKA-
gKA gKA gKA AKA gKA AKA gKA gKA AKA gKA gKA ggc ggA gCT gTc-
gKA gKA g-3' (SEQ ID NO:155)
```

5 Selección frente a codones de parada:

Debido a que algunas de las bibliotecas tienen codones NNK, tendrán algunos codones de parada TAG. Se podrían eliminar los codones con TAG clonando el ADN amplificado en un sitio *XbaI*-*BstEII* entre la secuencia señal para un gen *bla* y la proteína *bla* real, y expresándolo en células *Sup*⁰. Las colonias *Bla*^R no contienen paradas TAG. Como alternativa, se podrían clonar los fragmentos *XbaI*-*BstEII* delante de un gen de resistencia a canamicina y seleccionar para *Kan*^R. Entonces se movería el casete de *XbaI*-*BstEII* en la biblioteca de fago.

También, debido a que el titubeo permite algunos codones de parada, se puede mejorar la biblioteca eliminando los clones con paradas al clonar el ADN amplificado en un sitio *XbaI*-*BstEII* entre la secuencia señal para un gen *bla* y la proteína *bla* real, y expresarlo en células *Sup*⁰. Las colonias *Bla*^R no contienen paradas. Como alternativa, se pueden clonar los fragmentos *XbaI*-*BstEII* delante de un gen de resistencia a kanamicina, y seleccionar para *Kan*^R. Entonces se puede mover el casete de *XbaI*-*BstEII* en la biblioteca de fago.

Tabla 20: Regiones D humanas

!* para TAG; @ para TAA; \$ para TGA

D – Alineamiento de secuencias de aminoácidos (RF: marco de lectura)

		RF 1	RF 2	RF 3	Usados en diseños
D1	1-1	(SEQ ID NO:156) GITGT	(SEQ ID NO:157) VQLER	(SEQ ID NO:158) YNWNO	
	1-7	(SEQ ID NO:159) GITGT	(SEQ ID NO:160) V*LEL	(SEQ ID NO:161) YNWNY	
	1-20	(SEQ ID NO:159) GITGT	(SEQ ID NO:162) V*LER	(SEQ ID NO:163) YNWNO	
	1-26	(SEQ ID NO:164) GIVGAT	(SEQ ID NO:165) V*WELL	(SEQ ID NO:166) YSGSY	
D2	2-2	(SEQ ID NO:167) RIL**YQLLY	(SEQ ID NO:70) GYCSSTSCYT	(SEQ ID NO:168) DIVVVPAAI	1, 5, 6, 7,
	2-8	(SEQ ID NO:169) RILY@WCMLY	(SEQ ID NO:115) GYCTNGVCYT	(SEQ ID NO:170) DIVLMVYAI	20, 21, 22,
	2-15	(SEQ ID NO:171) RIL*WW*LLL	(SEQ ID NO:136) GYCSGGSCYS	(SEQ ID NO:172) DIVVVVAAT	25,
	2-21	(SEQ ID NO:173) SILW\$LLF	(SEQ ID NO:174) AYCGGDCYS	(SEQ ID NO:175) HIVVVTAI	
D3	3-3	(SEQ ID NO:176) VLRFLWLLY	(SEQ ID NO:177) YYDFWSGYYT	(SEQ ID NO:178) ITIFGVVVI	
	3-9	(SEQ ID NO:179) VLRYFDWLL@	(SEQ ID NO:180) YYDILTGYYN	(SEQ ID NO:181) ITIF*LVII	
	3-10	(SEQ ID NO:182) VLLWFGELL@	(SEQ ID NO:81) YYYGSYSYIN	(SEQ ID NO:183) ITMVRGVII	
	3-16	(SEQ ID NO:184) VL\$LRLEGELSLY	(SEQ ID NO:104) YYDYVWGSYRYT	(SEQ ID NO:185) IMITFGGVIVI	8, 9, 14, 15, 17, 18
	3-22	(SEQ ID NO:186) VLL***WLLL	(SEQ ID NO:187) YYDSSGYYY	(SEQ ID NO:188) ITMIVVVIT	4, 19, 20
D4	4-4	(SEQ ID NO:189) \$LQ@L	(SEQ ID NO:88) DYSNY	(SEQ ID NO:190) TTVT	
	4-11	(SEQ ID NO:191) \$LQ@L	(SEQ ID NO:192) DYSNY	(SEQ ID NO:193) TTVT	
	4-17	(SEQ ID NO:194) \$LR@L	(SEQ ID NO:195) DYGDY	(SEQ ID NO:196) TTVT	
	4-23	(SEQ ID NO:197) \$LRW@L	(SEQ ID NO:198) DYGNS	(SEQ ID NO:199) TTVVT	
D5	5-5	(SEQ ID NO:200) VDTAMV	(SEQ ID NO:201) WIQLWL	(SEQ ID NO:202) GYSGY	
	5-12	(SEQ ID NO:203) VDIVATI	(SEQ ID NO:204) WI*WLRL	(SEQ ID NO:205) GYSGYDY	
	5-18	(SEQ ID NO:206) VDTAMV	(SEQ ID NO:207) WIQLWL	(SEQ ID NO:208) GYSGY	
	5-24	(SEQ ID NO:209) VEMATI	(SEQ ID NO:210) *RWLQL	(SEQ ID NO:211) RDGYNY	
D6	6-6	(SEQ ID NO:212) EYSSSS	(SEQ ID NO:213) SIAAR	(SEQ ID NO:214) V*QLV	
	6-13	(SEQ ID NO:215) GYSSWY	(SEQ ID NO:216) GIAAAG	(SEQ ID NO:217) V*QQLV	
	6-19	(SEQ ID NO:218) GYSSGWY	(SEQ ID NO:219) GIAVAG	(SEQ ID NO:220) V*QWL	
D7	7-27	LTG	@LG	NWG	

Tabla 3: Segmentos de JH humano

JH – Alineamiento de secuencias de aminoácidos

	H3		

	CDR3		

	100	110	
		FR4-----	Usados en diseños
JH1	---AEYFOHWGQGLVTVSS	1-8,	(SEQ ID NO:66)
JH2	---YWYFDLWGRGTLVTVSS		(SEQ ID NO:67)
JH3	---AFDIWGQGTMTVTVSS		(SEQ ID NO:2)
JH4	---YFDYWGQGLVTVSS		(SEQ ID NO:1)
JH5	---NWFDPWGQGLVTVSS		(SEQ ID NO:68)
JH6	YYYYYGMDVWGQGTITVTVSS		(SEQ ID NO:3)
	123456789		

Tabla 10: ADN que codifica V-5D2-8.2a-JH2 para titubeo

```

!                                     CDR3.....
!   A   E   D   T   A   V   Y   Y   C   A   K   D   I   V   L   M
!   |gct|gag|gaT|aCT|GCA|GtT|ItaT|ItaC|tgc|gct  aag  jez  ezq  jzz  qzz  ezj
!
!   W   G   Q   G   T   T   V   T   V   S   S   (SEQ ID NO:224)
!   tgg  ggc  cag  ggt  act  acG  GTC  ACC  gtc  tcc  agt-3'   (SEQ ID NO:225)
!   BstEII...

```

Tabla 11: Trímeros que se pueden extraer de los segmentos D humanos

En las Tablas 11-14, el uso de una letra minúscula en una secuencia de aminoácidos indica que se cambió un codón de parada por el resto enunciado como la letra minúscula. Por ejemplo, en la secuencia de aminoácidos “yLE”, se introdujo un resto de Tyr en lugar de un codón de parada.

5

GTT	D1-1.1.1	1	
VQL	D1-1.2.1	2	
YNW	D1-1.3.1	3	
TTG	D1-1.1.2	4	
QLE	D1-1.2.2	5	
NWN	D1-1.3.2	6	
TGT	D1-1.1.3	7	
LER	D1-1.2.3	8	
WND	D1-1.3.3	9	
GIT	D1-7.1.1	10	
VyL	D1-7.2.1	11	*
ITG	D1-7.1.2	12	
yLE	D1-7.2.2	13	*
LEL	D1-7.2.3	14	
WNY	D1-7.3.3	15	
GIV	D1-26.1.1	16	
VyW	D1-26.2.1	17	*
YSG	D1-26.3.1	18	
IVG	D1-26.1.2	19	
yWE	D1-26.2.2	20	*
SGS	D1-26.3.2	21	

ES 2 528 963 T3

VGA	D1-26.1.3	22	
WEL	D1-26.2.3	23	
GSY	D1-26.3.3	24	
GAT	D1-26.1.4	25	
ELL	D1-26.2.4	26	
SY Y	D1-26.3.4	27	
RIL	D2-2.1.1	28	
GYC	D2-2.2.1	29	#
DIV	D2-2.3.1	30	
ILy	D2-2.1.2	31	*
YCS	D2-2.2.2	32	#
IVV	D2-2.3.2	33	
Lyy	D2-2.1.3	34	*
CSS	D2-2.2.3	35	#
VVV	D2-2.3.3	36	
yyY	D2-2.1.4	37	*
SST	D2-2.2.4	38	
VVP	D2-2.3.4	39	
yYQ	D2-2.1.5	40	*
STS	D2-2.2.5	41	
VPA	D2-2.3.5	42	
YQL	D2-2.1.6	43	
TSC	D2-2.2.6	44	#
PAA	D2-2.3.6	45	
QLL	D2-2.1.7	46	
SCY	D2-2.2.7	47	#
AAI	D2-2.3.7	48	
LLY	D2-2.1.8	49	
CYT	D2-2.2.8	50	#
ILY	D2-8.1.2	51	
YCT	D2-8.2.2	52	#
IVL	D2-8.3.2	53	
LYy	D2-8.1.3	54	*
CTN	D2-8.2.3	55	#
VLM	D2-8.3.3	56	
YyW	D2-8.1.4	57	*

ES 2 528 963 T3

TNG	D2-8.2.4	58	
LMV	D2-8.3.4	59	
yWC	D2-8.1.5	60	*#
NGV	D2-8.2.5	61	
MVY	D2-8.3.5	62	
WCM	D2-8.1.6	63	#
GVC	D2-8.2.6	64	#
VYA	D2-8.3.6	65	
CML	D2-8.1.7	66	#
VCY	D2-8.2.7	67	#
YAI	D2-8.3.7	68	
MLY	D2-8.1.8	69	
LyW	D2-15.1.3	70	*
CSG	D2-15.2.3	71	#
yWW	D2-15.1.4	72	*
SGG	D2-15.2.4	73	
WWy	D2-15.1.5	74	*
GGs	D2-15.2.5	75	
VVA	D2-15.3.5	76	
WyL	D2-15.1.6	77	*
GSC	D2-15.2.6	78	#
VAA	D2-15.3.6	79	
yLL	D2-15.1.7	80	*
AAT	D2-15.3.7	81	
LLL	D2-15.1.8	82	
CYS	D2-15.2.8	83	#
SIL	D2-21.1.1	84	
AYC	D2-21.2.1	85	#
HIV	D2-21.3.1	86	
ILW	D2-21.1.2	87	
YCG	D2-21.2.2	88	#
LWW	D2-21.1.3	89	
CGG	D2-21.2.3	90	#
WWw	D2-21.1.4	91	*
GGD	D2-21.2.4	92	
VVT	D2-21.3.4	93	

ES 2 528 963 T3

WwL	D2-21.1.5	94	*
GDC	D2-21.2.5	95	#
VTA	D2-21.3.5	96	
wLL	D2-21.1.6	97	*
DCY	D2-21.2.6	98	#
TAI	D2-21.3.6	99	
LLF	D2-21.1.7	100	
VLR	D3-3.1.1	101	
YYD	D3-3.2.1	102	
ITI	D3-3.3.1	103	
LRF	D3-3.1.2	104	
YDF	D3-3.2.2	105	
TIF	D3-3.3.2	106	
RFL	D3-3.1.3	107	
DFW	D3-3.2.3	108	
IFG	D3-3.3.3	109	
FLE	D3-3.1.4	110	
FWS	D3-3.2.4	111	
FGV	D3-3.3.4	112	
LEW	D3-3.1.5	113	
WSG	D3-3.2.5	114	
GVV	D3-3.3.5	115	
EWL	D3-3.1.6	116	
SGY	D3-3.2.6	117	
VVI	D3-3.3.6	118	
WLL	D3-3.1.7	119	
GYG	D3-3.2.7	120	
VII	D3-3.3.7	121	
YYT	D3-3.2.8	122	
LRY	D3-9.1.2	123	
YDI	D3-9.2.2	124	
RYF	D3-9.1.3	125	
DIL	D3-9.2.3	126	
IFy	D3-9.3.3	127	*
YFD	D3-9.1.4	128	
ILT	D3-9.2.4	129	

ES 2 528 963 T3

FyL	D3-9.3.4	130	*
FDW	D3-9.1.5	131	
LTG	D3-9.2.5	132	
yLV	D3-9.3.5	133	*
DWL	D3-9.1.6	134	
TGY	D3-9.2.6	135	
LVI	D3-9.3.6	136	
LLy	D3-9.1.8	137	*
YYN	D3-9.2.8	138	
VLL	D3-10.1.1	139	
YYY	D3-10.2.1	140	
ITM	D3-10.3.1	141	
LLW	D3-10.1.2	142	
YYG	D3-10.2.2	143	
TMV	D3-10.3.2	144	
LWF	D3-10.1.3	145	
YGS	D3-10.2.3	146	
MVR	D3-10.3.3	147	
WFG	D3-10.1.4	148	
GSG	D3-10.2.4	149	
VRG	D3-10.3.4	150	
FGE	D3-10.1.5	151	
RGV	D3-10.3.5	152	
GEL	D3-10.1.6	153	
GVI	D3-10.3.6	154	
VLw	D3-16.1.1	155	*
IMI	D3-16.3.1	156	
LwL	D3-16.1.2	157	*
YDY	D3-16.2.2	158	
MIT	D3-16.3.2	159	
wLR	D3-16.1.3	160	*
DYV	D3-16.2.3	161	
ITF	D3-16.3.3	162	
LRL	D3-16.1.4	163	
YVW	D3-16.2.4	164	
TFG	D3-16.3.4	165	

ES 2 528 963 T3

RLG	D3-16.1.5	166	
VWG	D3-16.2.5	167	
FGG	D3-16.3.5	168	
LGE	D3-16.1.6	169	
WGS	D3-16.2.6	170	
GGV	D3-16.3.6	171	
ELS	D3-16.1.8	172	
SYR	D3-16.2.8	173	
VIV	D3-16.3.8	174	
LSL	D3-16.1.9	175	
YRY	D3-16.2.9	176	
IVI	D3-16.3.9	177	
SLY	D3-16.1.10	178	
RYT	D3-16.2.10	179	
LLw	D3-22.1.2	180	*
TMI	D3-22.3.2	181	
Lwy	D3-22.1.3	182	*
YDS	D3-22.2.3	183	
MIV	D3-22.3.3	184	
wyy	D3-22.1.4	185	*
DSS	D3-22.2.4	186	
yyW	D3-22.1.5	187	*
SSG	D3-22.2.5	188	
yWL	D3-22.1.6	189	*
VIT	D3-22.3.7	190	
wLQ	D4-4.1.1	191	*
DYS	D4-4.2.1	192	
TTV	D4-4.3.1	193	
LQy	D4-4.1.2	194	*
YSN	D4-4.2.2	195	
TVT	D4-4.3.2	196	
QyL	D4-4.1.3	197	*
SNY	D4-4.2.3	198	
DYG	D4-17.2.1	199	
LRw	D4-17.1.2	200	*
YGD	D4-17.2.2	201	

ES 2 528 963 T3

RwL	D4-17.1.3	202	*
GDY	D4-17.2.3	203	
LRW	D4-23.1.2	204	
YGG	D4-23.2.2	205	
TVV	D4-23.3.2	206	
RWy	D4-23.1.3	207	*
GGN	D4-23.2.3	208	
GNS	D4-23.2.4	209	
VDT	D5-5.1.1	210	
WIQ	D5-5.2.1	211	
GYS	D5-5.3.1	212	
DTA	D5-5.1.2	213	
IQL	D5-5.2.2	214	
YSY	D5-5.3.2	215	
TAM	D5-5.1.3	216	
QLW	D5-5.2.3	217	
SYG	D5-5.3.3	218	
AMV	D5-5.1.4	219	
LWL	D5-5.2.4	220	
YGY	D5-5.3.4	221	
VDI	D5-12.1.1	222	
Wly	D5-12.2.1	223	*
lyW	D5-12.2.2	224	*
IVA	D5-12.1.3	225	
VAT	D5-12.1.4	226	
WLR	D5-12.2.4	227	
GYD	D5-12.3.4	228	
ATI	D5-12.1.5	229	
VEM	D5-24.1.1	230	
yRW	D5-24.2.1	231	*
RDG	D5-24.3.1	232	
EMA	D5-24.1.2	233	
RWL	D5-24.2.2	234	
DGY	D5-24.3.2	235	
MAT	D5-24.1.3	236	
WLQ	D5-24.2.3	237	

ES 2 528 963 T3

GYN	D5-24.3.3	238	
LQL	D5-24.2.4	239	
YNY	D5-24.3.4	240	
EYS	D6-6.1.1	241	
SIA	D6-6.2.1	242	
VyQ	D6-6.3.1	243	*
YSS	D6-6.1.2	244	
IAA	D6-6.2.2	245	
yQL	D6-6.3.2	246	*
SSS	D6-6.1.3	247	
AAR	D6-6.2.3	248	
QLV	D6-6.3.3	249	
GIA	D6-13.2.1	250	
yQQ	D6-13.3.2	251	*
AAA	D6-13.2.3	252	
QQL	D6-13.3.3	253	
SSW	D6-13.1.4	254	
AAG	D6-13.2.4	255	
SWY	D6-13.1.5	256	
IAV	D6-19.2.2	257	
yQW	D6-19.3.2	258	*
AVA	D6-19.2.3	259	
QWL	D6-19.3.3	260	
SGW	D6-19.1.4	261	
VAG	D6-19.2.4	262	
WLV	D6-19.3.4	263	
GWY	D6-19.1.5	264	
yLG	D7-27.2.1	265	*
NWG	D7-27.3.1	266	

Tabla 12: Tetrámeros distintos que se pueden extraer de segmentos D humanos

GTTG	D1-1.1.1	(SEQ ID NO:257)	1
VQLE	D1-1.2.1	(SEQ ID NO:258)	2
YNWN	D1-1.3.1	(SEQ ID NO:259)	3
TTGT	D1-1.1.2	(SEQ ID NO:263)	4
QLER	D1-1.2.2	(SEQ ID NO:264)	5

ES 2 528 963 T3

NWND	D1-1.3.2	(SEQ ID NO:265)	6
GITG	D1-7.1.1	(SEQ ID NO:266)	7
VyLE	D1-7.2.1	(SEQ ID NO:267)	8
ITGT	D1-7.1.2	(SEQ ID NO:271)	9
yLEL	D1-7.2.2	(SEQ ID NO:272)	10
NWNY	D1-7.3.2	(SEQ ID NO:273)	11
yLER	D1-20.2.2	(SEQ ID NO:275)	12
GIVG	D1-26.1.1	(SEQ ID NO:276)	13
VyWE	D1-26.2.1	(SEQ ID NO:277)	14
YSGS	D1-26.3.1	(SEQ ID NO:278)	15
IVGA	D1-26.1.2	(SEQ ID NO:285)	16
yWEL	D1-26.2.2	(SEQ ID NO:286)	17
SGSY	D1-26.3.2	(SEQ ID NO:287)	18
VGAT	D1-26.1.3	(SEQ ID NO:291)	19
WELL	D1-26.2.3	(SEQ ID NO:292)	20
GSYY	D1-26.3.3	(SEQ ID NO:293)	21
RILy	D2-2.1.1	(SEQ ID NO:294)	22
GYCS	D2-2.2.1	(SEQ ID NO:295)	23
DIWV	D2-2.3.1	(SEQ ID NO:296)	24
ILyy	D2-2.1.2	(SEQ ID NO:303)	25
YCSS	D2-2.2.2	(SEQ ID NO:304)	26
IVVV	D2-2.3.2	(SEQ ID NO:305)	27
LyyY	D2-2.1.3	(SEQ ID NO:312)	28
CSST	D2-2.2.3	(SEQ ID NO:313)	29
VVVP	D2-2.3.3	(SEQ ID NO:314)	30
yyYQ	D2-2.1.4	(SEQ ID NO:321)	31
SSTS	D2-2.2.4	(SEQ ID NO:322)	32
VVPA	D2-2.3.4	(SEQ ID NO:323)	33
yYQL	D2-2.1.5	(SEQ ID NO:330)	34
STSC	D2-2.2.5	(SEQ ID NO:331)	35
VPAA	D2-2.3.5	(SEQ ID NO:332)	36
YQLL	D2-2.1.6	(SEQ ID NO:338)	37
TSCY	D2-2.2.6	(SEQ ID NO:339)	38
PAAI	D2-2.3.6	(SEQ ID NO:340)	39
QLLY	D2-2.1.7	(SEQ ID NO:343)	40
SCYT	D2-2.2.7	(SEQ ID NO:344)	41

ES 2 528 963 T3

RILY	D2-8.1.1	(SEQ ID NO:345)	42
GYCT	D2-8.2.1	(SEQ ID NO:346)	43
DIVL	D2-8.3.1	(SEQ ID NO:347)	44
ILYy	D2-8.1.2	(SEQ ID NO:354)	45
YCTN	D2-8.2.2	(SEQ ID NO:355)	46
IVLM	D2-8.3.2	(SEQ ID NO:356)	47
LYyW	D2-8.1.3	(SEQ ID NO:363)	48
CTNG	D2-8.2.3	(SEQ ID NO:364)	49
VLMV	D2-8.3.3	(SEQ ID NO:365)	50
YyWC	D2-8.1.4	(SEQ ID NO:372)	51
TNGV	D2-8.2.4	(SEQ ID NO:373)	52
LMVY	D2-8.3.4	(SEQ ID NO:374)	53
yWCM	D2-8.1.5	(SEQ ID NO:381)	54
NGVC	D2-8.2.5	(SEQ ID NO:382)	55
MVYA	D2-8.3.5	(SEQ ID NO:383)	56
WCML	D2-8.1.6	(SEQ ID NO:389)	57
GVCY	D2-8.2.6	(SEQ ID NO:390)	58
VYAI	D2-8.3.6	(SEQ ID NO:391)	59
CMLY	D2-8.1.7	(SEQ ID NO:394)	60
VCYT	D2-8.2.7	(SEQ ID NO:395)	61
ILyW	D2-15.1.2	(SEQ ID NO:401)	62
YCSG	D2-15.2.2	(SEQ ID NO:402)	63
LyWW	D2-15.1.3	(SEQ ID NO:409)	64
CSGG	D2-15.2.3	(SEQ ID NO:410)	65
VVVV	D2-15.3.3	(SEQ ID NO:411)	66
yWWy	D2-15.1.4	(SEQ ID NO:418)	67
SGGS	D2-15.2.4	(SEQ ID NO:419)	68
VVVA	D2-15.3.4	(SEQ ID NO:420)	69
WWyL	D2-15.1.5	(SEQ ID NO:427)	70
GGSC	D2-15.2.5	(SEQ ID NO:428)	71
VVAA	D2-15.3.5	(SEQ ID NO:429)	72
WyLL	D2-15.1.6	(SEQ ID NO:435)	73
GSCY	D2-15.2.6	(SEQ ID NO:436)	74
VAAT	D2-15.3.6	(SEQ ID NO:437)	75
yLLL	D2-15.1.7	(SEQ ID NO:440)	76
SCYS	D2-15.2.7	(SEQ ID NO:441)	77

ES 2 528 963 T3

SILW	D2-21.1.1	(SEQ ID NO:442)	78
AYCG	D2-21.2.1	(SEQ ID NO:443)	79
HIVV	D2-21.3.1	(SEQ ID NO:444)	80
ILWW	D2-21.1.2	(SEQ ID NO:451)	81
YCGG	D2-21.2.2	(SEQ ID NO:452)	82
LWWw	D2-21.1.3	(SEQ ID NO:459)	83
CGGD	D2-21.2.3	(SEQ ID NO:460)	84
VVVT	D2-21.3.3	(SEQ ID NO:461)	85
WWwL	D2-21.1.4	(SEQ ID NO:468)	86
GGDC	D2-21.2.4	(SEQ ID NO:469)	87
VVTA	D2-21.3.4	(SEQ ID NO:470)	88
WwLL	D2-21.1.5	(SEQ ID NO:476)	89
GDCY	D2-21.2.5	(SEQ ID NO:477)	90
VTAI	D2-21.3.5	(SEQ ID NO:478)	91
wLLF	D2-21.1.6	(SEQ ID NO:481)	92
DCYS	D2-21.2.6	(SEQ ID NO:482)	93
VLRF	D3-3.1.1	(SEQ ID NO:483)	94
YYDF	D3-3.2.1	(SEQ ID NO:484)	95
ITIF	D3-3.3.1	(SEQ ID NO:485)	96
LRFL	D3-3.1.2	(SEQ ID NO:492)	97
YDFW	D3-3.2.2	(SEQ ID NO:493)	98
TIFG	D3-3.3.2	(SEQ ID NO:494)	99
RFLE	D3-3.1.3	(SEQ ID NO:501)	100
DFWS	D3-3.2.3	(SEQ ID NO:502)	101
IFGV	D3-3.3.3	(SEQ ID NO:503)	102
FLEW	D3-3.1.4	(SEQ ID NO:510)	103
FWSG	D3-3.2.4	(SEQ ID NO:511)	104
FGVV	D3-3.3.4	(SEQ ID NO:512)	105
LEWL	D3-3.1.5	(SEQ ID NO:519)	106
WSGY	D3-3.2.5	(SEQ ID NO:520)	107
GVVI	D3-3.3.5	(SEQ ID NO:521)	108
EWLL	D3-3.1.6	(SEQ ID NO:527)	109
SGYY	D3-3.2.6	(SEQ ID NO:528)	110
VVII	D3-3.3.6	(SEQ ID NO:529)	111
WLLY	D3-3.1.7	(SEQ ID NO:532)	112
GYYT	D3-3.2.7	(SEQ ID NO:533)	113

ES 2 528 963 T3

VLRY	D3-9.1.1	(SEQ ID NO:534)	114
YYDI	D3-9.2.1	(SEQ ID NO:535)	115
LRYF	D3-9.1.2	(SEQ ID NO:542)	116
YDIL	D3-9.2.2	(SEQ ID NO:543)	117
TIFy	D3-9.3.2	(SEQ ID NO:544)	118
RYFD	D3-9.1.3	(SEQ ID NO:551)	119
DILT	D3-9.2.3	(SEQ ID NO:552)	120
IFyL	D3-9.3.3	(SEQ ID NO:553)	121
YFDW	D3-9.1.4	(SEQ ID NO:560)	122
ILTG	D3-9.2.4	(SEQ ID NO:561)	123
FyLV	D3-9.3.4	(SEQ ID NO:562)	124
FDWL	D3-9.1.5	(SEQ ID NO:569)	125
LTGY	D3-9.2.5	(SEQ ID NO:570)	126
yLVI	D3-9.3.5	(SEQ ID NO:571)	127
DWLL	D3-9.1.6	(SEQ ID NO:577)	128
TGYY	D3-9.2.6	(SEQ ID NO:578)	129
LVII	D3-9.3.6	(SEQ ID NO:579)	130
WLLy	D3-9.1.7	(SEQ ID NO:582)	131
GYYN	D3-9.2.7	(SEQ ID NO:583)	132
VLLW	D3-10.1.1	(SEQ ID NO:584)	133
YYYG	D3-10.2.1	(SEQ ID NO:585)	134
ITMV	D3-10.3.1	(SEQ ID NO:586)	135
LLWF	D3-10.1.2	(SEQ ID NO:593)	136
YYGS	D3-10.2.2	(SEQ ID NO:594)	137
TMVR	D3-10.3.2	(SEQ ID NO:595)	138
LWFG	D3-10.1.3	(SEQ ID NO:602)	139
YGSG	D3-10.2.3	(SEQ ID NO:603)	140
MVRG	D3-10.3.3	(SEQ ID NO:604)	141
WFGE	D3-10.1.4	(SEQ ID NO:611)	142
GSGS	D3-10.2.4	(SEQ ID NO:612)	143
VRGV	D3-10.3.4	(SEQ ID NO:613)	144
FGEL	D3-10.1.5	(SEQ ID NO:620)	145
RGVI	D3-10.3.5	(SEQ ID NO:621)	146
GELL	D3-10.1.6	(SEQ ID NO:626)	147
GVII	D3-10.3.6	(SEQ ID NO:627)	148
ELLy	D3-10.1.7	(SEQ ID NO:630)	149

ES 2 528 963 T3

SYYN	D3-10.2.7	(SEQ ID NO:631)	150
VLwL	D3-16.1.1	(SEQ ID NO:632)	151
YYDY	D3-16.2.1	(SEQ ID NO:633)	152
IMIT	D3-16.3.1	(SEQ ID NO:634)	153
LwLR	D3-16.1.2	(SEQ ID NO:641)	154
YDYV	D3-16.2.2	(SEQ ID NO:642)	155
MITF	D3-16.3.2	(SEQ ID NO:643)	156
wLRL	D3-16.1.3	(SEQ ID NO:650)	157
DYVW	D3-16.2.3	(SEQ ID NO:651)	158
ITFG	D3-16.3.3	(SEQ ID NO:652)	159
LRLG	D3-16.1.4	(SEQ ID NO:659)	160
YVWG	D3-16.2.4	(SEQ ID NO:660)	161
TFGG	D3-16.3.4	(SEQ ID NO:661)	162
RLGE	D3-16.1.5	(SEQ ID NO:668)	163
VWGS	D3-16.2.5	(SEQ ID NO:669)	164
FGGV	D3-16.3.5	(SEQ ID NO:670)	165
LGEL	D3-16.1.6	(SEQ ID NO:677)	166
WGSY	D3-16.2.6	(SEQ ID NO:678)	167
GGVI	D3-16.3.6	(SEQ ID NO:679)	168
GELS	D3-16.1.7	(SEQ ID NO:686)	169
GSYR	D3-16.2.7	(SEQ ID NO:687)	170
GVIV	D3-16.3.7	(SEQ ID NO:688)	171
ELSL	D3-16.1.8	(SEQ ID NO:694)	172
SYRY	D3-16.2.8	(SEQ ID NO:695)	173
VIVI	D3-16.3.8	(SEQ ID NO:696)	174
LSLY	03-16.1.9	(SEQ ID NO:699)	175
YRYT	D3-16.2.9	(SEQ ID NO:700)	176
VLLw	D3-22.1.1	(SEQ ID NO:701)	177
YYYD	D3-22.2.1	(SEQ ID NO:702)	178
ITMI	D3-22.3.1	(SEQ ID NO:703)	179
LLwy	D3-22.1.2	(SEQ ID NO:710)	180
YYDS	D3-22.2.2	(SEQ ID NO:711)	181
TMIV	D3-22.3.2	(SEQ ID NO:712)	182
Lwyy	D3-22.1.3	(SEQ ID NO:719)	183
YDSS	D3-22.2.3	(SEQ ID NO:720)	184
MIVV	D3-22.3.3	(SEQ ID NO:721)	185

ES 2 528 963 T3

wyyW	D3-22.1.4	(SEQ ID NO:728)	186
DSSG	D3-22.2.4	(SEQ ID NO:729)	187
yyWL	D3-22.1.5	(SEQ ID NO:736)	188
SSGY	D3-22.2.5	(SEQ ID NO:737)	189
VVVI	D3-22.3.5	(SEQ ID NO:738)	190
yWLL	D3-22.1.6	(SEQ ID NO:744)	191
VVIT	D3-22.3.6	(SEQ ID NO:745)	192
WLLL	D3-22.1.7	(SEQ ID NO:748)	193
GYYY	D3-22.2.7	(SEQ ID NO:749)	194
wLQy	D4-4.1.1	(SEQ ID NO:750)	195
DYSN	D4-4.2.1	(SEQ ID NO:751)	196
TTVT	D4-4.3.1	(SEQ ID NO:752)	197
LQyL	D4-4.1.2	(SEQ ID NO:755)	198
YSNY	D4-4.2.2	(SEQ ID NO:756)	199
wLRw	D4-17.1.1	(SEQ ID NO:757)	200
DYGD	D4-17.2.1	(SEQ ID NO:758)	201
LRwL	D4-17.1.2	(SEQ ID NO:761)	202
YGDY	D4-17.2.2	(SEQ ID NO:762)	203
wLRW	D4-23.1.1	(SEQ ID NO:763)	204
DYGG	D4-23.2.1	(SEQ ID NO:764)	205
TTVV	D4-23.3.1	(SEQ ID NO:765)	206
LRWy	D4-23.1.2	(SEQ ID NO:771)	207
YGGN	D4-23.2.2	(SEQ ID NO:772)	208
TVVT	D4-23.3.2	(SEQ ID NO:773)	209
RWyL	D4-23.1.3	(SEQ ID NO:776)	210
GGNS	D4-23.2.3	(SEQ ID NO:777)	211
VDTA	D5-5.1.1	(SEQ ID NO:778)	212
WIQL	D5-5.2.1	(SEQ ID NO:779)	213
GYSY	D5-5.3.1	(SEQ ID NO:780)	214
DTAM	D5-5.1.2	(SEQ ID NO:787)	215
IQLW	D5-5.2.2	(SEQ ID NO:788)	216
YSYG	D5-5.3.2	(SEQ ID NO:789)	217
TAMV	D5-5.1.3	(SEQ ID NO:793)	218
QLWL	D5-5.2.3	(SEQ ID NO:794)	219
SYGY	D5-5.3.3	(SEQ ID NO:795)	220
VDIV	D5-12.1.1	(SEQ ID NO:796)	221

ES 2 528 963 T3

WlyW	D5-12.2.1	(SEQ ID NO:797)	222
GYSG	D5-12.3.1	(SEQ ID NO:798)	223
DIVA	D5-12.1.2	(SEQ ID NO:805)	224
lyWL	D5-12.2.2	(SEQ ID NO:806)	225
YSGY	D5-12.3.2	(SEQ ID NO:807)	226
IVAT	D5-12.1.3	(SEQ ID NO:814)	227
yWLR	D5-12.2.3	(SEQ ID NO:815)	228
SGYD	D5-12.3.3	(SEQ ID NO:816)	229
VATI	D5-12.1.4	(SEQ ID NO:820)	230
WLRL	D5-12.2.4	(SEQ ID NO:821)	231
GYDY	D5-12.3.4	(SEQ ID NO:822)	232
VEMA	D5-24.1.1	(SEQ ID NO:823)	233
yRWL	D5-24.2.1	(SEQ ID NO:824)	234
RDGY	D5-24.3.1	(SEQ ID NO:825)	235
EMAT	D5-24.1.2	(SEQ ID NO:832)	236
RWLQ	D5-24.2.2	(SEQ ID NO:833)	237
DGYN	D5-24.3.2	(SEQ ID NO:834)	238
MATI	D5-24.1.3	(SEQ ID NO:838)	239
WLQL	D5-24.2.3	(SEQ ID NO:839)	240
GYNY	D5-24.3.3	(SEQ ID NO:840)	241
EYSS	D6-6.1.1	(SEQ ID NO:841)	242
SIAA	D6-6.2.1	(SEQ ID NO:842)	243
VyQL	D6-6.3.1	(SEQ ID NO:843)	244
YSSS	D6-6.1.2	(SEQ ID NO:848)	245
IAAR	D6-6.2.2	(SEQ ID NO:849)	246
yQLV	D6-6.3.2	(SEQ ID NO:850)	247
SSSS	D6-6.1.3	(SEQ ID NO:852)	248
GYSS	D6-13.1.1	(SEQ ID NO:853)	249
GIAA	D6-13.2.1	(SEQ ID NO:854)	250
VyQQ	D6-13.3.1	(SEQ ID NO:855)	251
IAAA	D6-13.2.2	(SEQ ID NO:862)	252
yQQL	D6-13.3.2	(SEQ ID NO:863)	253
SSSW	D6-13.1.3	(SEQ ID NO:868)	254
AAAG	D6-13.2.3	(SEQ ID NO:869)	255
QQLV	D6-13.3.3	(SEQ ID NO:870)	256
SSWY	D6-13.1.4	(SEQ ID NO:872)	257

GIAV	D6-19.2.1	(SEQ ID NO:873)	258
VyQW	D6-19.3.1	(SEQ ID NO:874)	259
YSSG	D6-19.1.2	(SEQ ID NO:881)	260
IAVA	D6-19.2.2	(SEQ ID NO:882)	261
yQWL	D6-19.3.2	(SEQ ID NO:883)	262
SSGW	D6-19.1.3	(SEQ ID NO:888)	263
AVAG	D6-19.2.3	(SEQ ID NO:889)	264
QWLV	D6-19.3.3	(SEQ ID NO:890)	265
SGWY	D6-19.1.4	(SEQ ID NO:892)	266

Tabla 13: Pentámeros que se pueden extraer de segmentos D humanos

GTTGT	D1-1.1.1	(SEQ ID NO:260)	1
VQLER	D1-1.2.1	(SEQ ID NO:261)	2
YNWND	D1-1.3.1	(SEQ ID NO:262)	3
GITGT	D1-7.1.1	(SEQ ID NO:268)	4
VyLEL	D1-7.2.1	(SEQ ID NO:269)	5
YNWNY	D1-7.3.1	(SEQ ID NO:270)	6
VyLER	D1-20.2.1	(SEQ ID NO:274)	7
GIVGA	D1-26.1.1	(SEQ ID NO:279)	8
VyWEL	D1-26.2.1	(SEQ ID NO:280)	9
YSGSY	D1-26.3.1	(SEQ ID NO:281)	10
IVGAT	D1-26.1.2	(SEQ ID NO:288)	11
yWELL	D1-26.2.2	(SEQ ID NO:289)	12
SGSYY	D1-26.3.2	(SEQ ID NO:290)	13
RIlyy	D2-2.1.1	(SEQ ID NO:297)	14
GYCSS	D2-2.2.1	(SEQ ID NO:298)	15
DIVVV	D2-2.3.1	(SEQ ID NO:299)	16
ILyyY	D2-2.1.2	(SEQ ID NO:306)	17
YCSST	D2-2.2.2	(SEQ ID NO:307)	18
IVVVP	D2-2.3.2	(SEQ ID NO:308)	19
LyyYQ	D2-2.1.3	(SEQ ID NO:315)	20
CSSTS	D2-2.2.3	(SEQ ID NO:316)	21
VVPPA	D2-2.3.3	(SEQ ID NO:317)	22
yyYQL	D2-2.1.4	(SEQ ID NO:324)	23
SSTSC	D2-2.2.4	(SEQ ID NO:325)	24
VVPAA	D2-2.3.4	(SEQ ID NO:326)	25

ES 2 528 963 T3

yYQLL	D2-2.1.5	(SEQ ID NO:333)	26
STSCY	D2-2.2.5	(SEQ ID NO:334)	27
VPAAI	D2-2.3.5	(SEQ ID NO:335)	28
YQLLY	D2-2.1.6	(SEQ ID NO:341)	29
TSCYT	D2-2.2.6	(SEQ ID NO:342)	30
RILYy	D2-8.1.1	(SEQ ID NO:348)	31
GYCTN	D2-8.2.1	(SEQ ID NO:349)	32
DIVLM	D2-8.3.1	(SEQ ID NO:350)	33
ILYyW	D2-8.1.2	(SEQ ID NO:357)	34
YCTNG	D2-8.2.2	(SEQ ID NO:358)	35
IVLMV	D2-8.3.2	(SEQ ID NO:359)	36
LYyWC	D2-8.1.3	(SEQ ID NO:366)	37
CTNGV	D2-8.2.3	(SEQ ID NO:367)	38
VLMVY	D2-8.3.3	(SEQ ID NO:368)	39
YyWCM	D2-8.1.4	(SEQ ID NO:375)	40
TNGVC	D2-8.2.4	(SEQ ID NO:376)	41
LMVYA	D2-8.3.4	(SEQ ID NO:377)	42
yWCML	D2-8.1.5	(SEQ ID NO:384)	43
NGVCY	D2-8.2.5	(SEQ ID NO:385)	44
MVYAI	D2-8.3.5	(SEQ ID NO:386)	45
WCMLY	D2-8.1.6	(SEQ ID NO:392)	46
GVCYT	D2-8.2.6	(SEQ ID NO:393)	47
RILyW	D2-15.1.1	(SEQ ID NO:396)	48
GYCSG	D2-15.2.1	(SEQ ID NO:397)	49
ILyWW	D2-15.1.2	(SEQ ID NO:403)	50
YCSGG	D2-15.2.2	(SEQ ID NO:404)	51
IVVVV	D2-15.3.2	(SEQ ID NO:405)	52
LyWWy	D2-15.1.3	(SEQ ID NO:412)	53
CSGGS	D2-15.2.3	(SEQ ID NO:413)	54
VVVVA	D2-15.3.3	(SEQ ID NO:414)	55
yWWyL	D2-15.1.4	(SEQ ID NO:421)	56
SGGSC	D2-15.2.4	(SEQ ID NO:422)	57
VVVAA	D2-15.3.4	(SEQ ID NO:423)	58
WWyLL	D2-15.1.5	(SEQ ID NO:430)	59
GGSCY	D2-15.2.5	(SEQ ID NO:431)	60
VVAAT	D2-15.3.5	(SEQ ID NO:432)	61

ES 2 528 963 T3

WyLLL	D2-15.1.6	(SEQ ID NO:438)	62
GSCYS	D2-15.2.6	(SEQ ID NO:439)	63
SILWW	D2-21.1.1	(SEQ ID NO:445)	64
AYCGG	D2-21.2.1	(SEQ ID NO:446)	65
HIVVV	D2-21.3.1	(SEQ ID NO:447)	66
ILWWw	D2-21.1.2	(SEQ ID NO:453)	67
YCGGD	D2-21.2.2	(SEQ ID NO:454)	68
IVVVT	D2-21.3.2	(SEQ ID NO:455)	69
LWWwL	D2-21.1.3	(SEQ ID NO:462)	70
CGGDC	D2-21.2.3	(SEQ ID NO:463)	71
VVVTa	D2-21.3.3	(SEQ ID NO:464)	72
WWwLL	D2-21.1.4	(SEQ ID NO:471)	73
GGDCY	D2-21.2.4	(SEQ ID NO:472)	74
VVTAI	D2-21.3.4	(SEQ ID NO:473)	75
WwLLF	D2-21.1.5	(SEQ ID NO:479)	76
GDCYS	D2-21.2.5	(SEQ ID NO:480)	77
VLRFL	D3-3.1.1	(SEQ ID NO:486)	78
YYDFW	D3-3.2.1	(SEQ ID NO:487)	79
ITIFG	D3-3.3.1	(SEQ ID NO:488)	80
LRFLE	D3-3.1.2	(SEQ ID NO:495)	81
YDFWS	D3-3.2.2	(SEQ ID NO:496)	82
TIFGV	D3-3.3.2	(SEQ ID NO:497)	83
RFLEW	D3-3.1.3	(SEQ ID NO:504)	84
DFWSG	D3-3.2.3	(SEQ ID NO:505)	85
IFGVV	D3-3.3.3	(SEQ ID NO:506)	86
FLEWL	D3-3.1.4	(SEQ ID NO:513)	87
FWSGY	D3-3.2.4	(SEQ ID NO:514)	88
FGVVI	D3-3.3.4	(SEQ ID NO:515)	89
LEWLL	D3-3.1.5	(SEQ ID NO:522)	90
WSGYY	D3-3.2.5	(SEQ ID NO:523)	91
GVVII	D3-3.3.5	(SEQ ID NO:524)	92
EWLLY	D3-3.1.6	(SEQ ID NO:530)	93
SGYYT	D3-3.2.6	(SEQ ID NO:531)	94
VLRYF	D3-9.1.1	(SEQ ID NO:536)	95
YYDIL	D3-9.2.1	(SEQ ID NO:537)	96
ITIFy	D3-9.3.1	(SEQ ID NO:538)	97

ES 2 528 963 T3

LRYFD	D3-9.1.2	(SEQ ID NO:545)	98
YDILT	D3-9.2.2	(SEQ ID NO:546)	99
TIFyL	D3-9.3.2	(SEQ ID NO:547)	100
RYFDW	D3-9.1.3	(SEQ ID NO:554)	101
DILTG	D3-9.2.3	(SEQ ID NO:555)	102
IFyLV	D3-9.3.3	(SEQ ID NO:556)	103
YFDWL	D3-9.1.4	(SEQ ID NO:563)	104
ILTGY	D3-9.2.4	(SEQ ID NO:564)	105
FyLVI	D3-9.3.4	(SEQ ID NO:565)	106
FDWLL	D3-9.1.5	(SEQ ID NO:572)	107
LTGYY	D3-9.2.5	(SEQ ID NO:573)	108
yLVII	D3-9.3.5	(SEQ ID NO:574)	109
DWLLy	D3-9.1.6	(SEQ ID NO:580)	110
TGYYN	D3-9.2.6	(SEQ ID NO:581)	111
VLLWF	D3-10.1.1	(SEQ ID NO:587)	112
YYYGS	D3-10.2.1	(SEQ ID NO:588)	113
ITMVR	D3-10.3.1	(SEQ ID NO:589)	114
LLWFG	D3-10.1.2	(SEQ ID NO:596)	115
YYGSG	D3-10.2.2	(SEQ ID NO:597)	116
TMVRG	D3-10.3.2	(SEQ ID NO:598)	117
LWFGE	D3-10.1.3	(SEQ ID NO:605)	118
YGSGS	D3-10.2.3	(SEQ ID NO:606)	119
MVRGV	D3-10.3.3	(SEQ ID NO:607)	120
WFGEL	D3-10.1.4	(SEQ ID NO:614)	121
GSGSY	D3-10.2.4	(SEQ ID NO:615)	122
VRGVI	D3-10.3.4	(SEQ ID NO:616)	123
FGELL	D3-10.1.5	(SEQ ID NO:622)	124
RGVII	D3-10.3.5	(SEQ ID NO:623)	125
GELLy	D3-10.1.6	(SEQ ID NO:628)	126
GSYYN	D3-10.2.6	(SEQ ID NO:629)	127
VLwLR	D3-16.1.1	(SEQ ID NO:635)	128
YYDYV	D3-16.2.1	(SEQ ID NO:636)	129
IMITF	D3-16.3.1	(SEQ ID NO:637)	130
LwLRL	D3-16.1.2	(SEQ ID NO:644)	131
YDYVW	D3-16.2.2	(SEQ ID NO:645)	132
MITFG	D3-16.3.2	(SEQ ID NO:646)	133

ES 2 528 963 T3

wLRLG	D3-16.1.3	(SEQ ID NO:653)	134
DYVWG	D3-16.2.3	(SEQ ID NO:654)	135
ITFGG	D3-16.3.3	(SEQ ID NO:655)	136
LRLGE	D3-16.1.4	(SEQ ID NO:662)	137
YVWGS	D3-16.2.4	(SEQ ID NO:663)	138
TFGGV	D3-16.3.4	(SEQ ID NO:664)	139
RLGEL	D3-16.1.5	(SEQ ID NO:671)	140
VWGSY	D3-16.2.5	(SEQ ID NO:672)	141
FGGVI	D3-16.3.5	(SEQ ID NO:673)	142
LGELS	D3-16.1.6	(SEQ ID NO:680)	143
WGSYR	D3-16.2.6	(SEQ ID NO:681)	144
GGVIV	D3-16.3.6	(SEQ ID NO:682)	145
GELSL	D3-16.1.7	(SEQ ID NO:689)	146
GSYRY	D3-16.2.7	(SEQ ID NO:690)	147
GVIVI	D3-16.3.7	(SEQ ID NO:691)	148
ELSLY	D3-16.1.8	(SEQ ID NO:697)	149
SYRYT	D3-16.2.8	(SEQ ID NO:698)	150
VLLwy	D3-22.1.1	(SEQ ID NO:704)	151
YYYDS	D3-22.2.1	(SEQ ID NO:705)	152
ITMIV	D3-22.3.1	(SEQ ID NO:706)	153
LLwyy	D3-22.1.2	(SEQ ID NO:713)	154
YYDSS	D3-22.2.2	(SEQ ID NO:714)	155
TMIVV	D3-22.3.2	(SEQ ID NO:715)	156
LwyyW	D3-22.1.3	(SEQ ID NO:722)	157
YDSSG	D3-22.2.3	(SEQ ID NO:723)	158
MIVVV	D3-22.3.3	(SEQ ID NO:724)	159
wyyWL	D3-22.1.4	(SEQ ID NO:730)	160
DSSGY	D3-22.2.4	(SEQ ID NO:731)	161
IVVVI	D3-22.3.4	(SEQ ID NO:732)	162
yyWLL	D3-22.1.5	(SEQ ID NO:739)	163
SSGY Y	D3-22.2.5	(SEQ ID NO:740)	164
VVVIT	D3-22.3.5	(SEQ ID NO:741)	165
yWLLL	D3-22.1.6	(SEQ ID NO:746)	166
SGYYY	D3-22.2.6	(SEQ ID NO:747)	167
wLQyL	D4-4.1.1	(SEQ ID NO:753)	168
DYSNY	D4-4.2.1	(SEQ ID NO:754)	169

ES 2 528 963 T3

wLRwL	D4-17.1.1	(SEQ ID NO:759)	170
DYGDY	D4-17.2.1	(SEQ ID NO:760)	171
wLRWy	D4-23.1.1	(SEQ ID NO:766)	172
DYGGN	D4-23.2.1	(SEQ ID NO:767)	173
TTVVT	D4-23.3.1	(SEQ ID NO:768)	174
LRWyL	D4-23.1.2	(SEQ ID NO:774)	175
YGGNS	D4-23.2.2	(SEQ ID NO:775)	176
VDTAM	D5-5.1.1	(SEQ ID NO:781)	177
WIQLW	D5-5.2.1	(SEQ ID NO:782)	178
GYSYG	D5-5.3.1	(SEQ ID NO:783)	179
DTAMV	D5-5.1.2	(SEQ ID NO:790)	180
IQLWL	D5-5.2.2	(SEQ ID NO:791)	181
YSYGY	D5-5.3.2	(SEQ ID NO:792)	182
VDIVA	D5-12.1.1	(SEQ ID NO:799)	183
WlyWL	D5-12.2.1	(SEQ ID NO:800)	184
GYSGY	D5-12.3.1	(SEQ ID NO:801)	185
DIVAT	D5-12.1.2	(SEQ ID NO:808)	186
lyWLR	D5-12.2.2	(SEQ ID NO:809)	187
YSGYD	D5-12.3.2	(SEQ ID NO:810)	188
IVATI	D5-12.1.3	(SEQ ID NO:817)	189
yWLRL	D5-12.2.3	(SEQ ID NO:818)	190
SGYDY	D5-12.3.3	(SEQ ID NO:819)	191
VEMAT	D5-24.1.1	(SEQ ID NO:826)	192
yRWLQ	D5-24.2.1	(SEQ ID NO:827)	193
RDGYN	D5-24.3.1	(SEQ ID NO:828)	194
EMATI	D5-24.1.2	(SEQ ID NO:835)	195
RWLQL	D5-24.2.2	(SEQ ID NO:836)	196
DGYNY	D5-24.3.2	(SEQ ID NO:837)	197
EYSSS	D6-6.1.1	(SEQ ID NO:844)	198
SIAAR	D6-6.2.1	(SEQ ID NO:845)	199
VyQLV	D6-6.3.1	(SEQ ID NO:846)	200
YSSSS	D6-6.1.2	(SEQ ID NO:851)	201
GYSSS	D6-13.1.1	(SEQ ID NO:856)	202
GIAAA	D6-13.2.1	(SEQ ID NO:857)	203
VyQQL	D6-13.3.1	(SEQ ID NO:858)	204
YSSSW	D6-13.1.2	(SEQ ID NO:864)	205

IAAAG	D6-13.2.2	(SEQ ID NO:865)	206
yQQLV	D6-13.3.2	(SEQ ID NO:866)	207
SSSWY	D6-13.1.3	(SEQ ID NO:871)	208
GYSSG	D6-19.1.1	(SEQ ID NO:875)	209
GIAVA	D6-19.2.1	(SEQ ID NO:876)	210
VyQWL	D6-19.3.1	(SEQ ID NO:877)	211
YSSGW	D6-19.1.2	(SEQ ID NO:884)	212
IAVAG	D6-19.2.2	(SEQ ID NO:885)	213
yQWLV	D6-19.3.2	(SEQ ID NO:886)	214
SSGWY	D6-19.1.3	(SEQ ID NO:891)	215

Tabla 14: Todos los hexámeros que se pueden extraer de segmentos D humanos

GIVGAT	D1-26.1.1	(SEQ ID NO:282)	1
VyWELL	D1-26.2.1	(SEQ ID NO:283)	2
YSGSYY	D1-26.3.1	(SEQ ID NO:284)	3
RILyyY	D2-2.1.1	(SEQ ID NO:300)	4
GYCSST	D2-2.2.1	(SEQ ID NO:301)	5
DIVVVP	D2-2.3.1	(SEQ ID NO:302)	6
ILyyYQ	D2-2.1.2	(SEQ ID NO:309)	7
YCSSTS	D2-2.2.2	(SEQ ID NO:310)	8
IVVPA	D2-2.3.2	(SEQ ID NO:311)	9
LyyYQL	D2-2.1.3	(SEQ ID NO:318)	10
CSSTSC	D2-2.2.3	(SEQ ID NO:319)	11
VVPAA	D2-2.3.3	(SEQ ID NO:320)	12
yyYQLL	D2-2.1.4	(SEQ ID NO:327)	13
SSTSCY	D2-2.2.4	(SEQ ID NO:328)	14
VVPAAI	D2-2.3.4	(SEQ ID NO:329)	15
yYQLLY	D2-2.1.5	(SEQ ID NO:336)	16
STSCYT	D2-2.2.5	(SEQ ID NO:337)	17
RILYyW	D2-8.1.1	(SEQ ID NO:351)	18
GYCTNG	D2-8.2.1	(SEQ ID NO:352)	19
DIVLMV	D2-8.3.1	(SEQ ID NO:353)	20
ILYyWC	D2-8.1.2	(SEQ ID NO:360)	21
YCTNGV	D2-8.2.2	(SEQ ID NO:361)	22
IVLMVY	D2-8.3.2	(SEQ ID NO:362)	23
LYyWCM	D2-8.1.3	(SEQ ID NO:369)	24

ES 2 528 963 T3

CTNGVC	D2-8.2.3	(SEQ ID NO:370)	25
VLMVYA	D2-8.3.3	(SEQ ID NO:371)	26
YyWCML	D2-8.1.4	(SEQ ID NO:378)	27
TNGVCY	D2-8.2.4	(SEQ ID NO:379)	28
LMVYAI	D2-8.3.4	(SEQ ID NO:380)	29
yWCMLY	D2-8.1.5	(SEQ ID NO:387)	30
NGVCYT	D2-8.2.5	(SEQ ID NO:388)	31
RILyWW	D2-15.1.1	(SEQ ID NO:398)	32
GYCSGG	D2-15.2.1	(SEQ ID NO:399)	33
DIVVVV	D2-15.3.1	(SEQ ID NO:400)	34
ILyWWy	D2-15.1.2	(SEQ ID NO:406)	35
YCSGGS	D2-15.2.2	(SEQ ID NO:407)	36
IVVVA	D2-15.3.2	(SEQ ID NO:408)	37
LyWWyL	D2-15.1.3	(SEQ ID NO:415)	38
CSGGSC	D2-15.2.3	(SEQ ID NO:416)	39
VVVVAA	D2-15.3.3	(SEQ ID NO:417)	40
yWWyLL	D2-15.1.4	(SEQ ID NO:424)	41
SGGSCY	D2-15.2.4	(SEQ ID NO:425)	42
VVVAAT	D2-15.3.4	(SEQ ID NO:426)	43
WWyLLL	D2-15.1.5	(SEQ ID NO:433)	44
GGSCYS	D2-15.2.5	(SEQ ID NO:434)	45
SILWWw	D2-21.1.1	(SEQ ID NO:448)	46
AYCGGD	D2-21.2.1	(SEQ ID NO:449)	47
HIVVVT	D2-21.3.1	(SEQ ID NO:450)	48
ILWWwL	D2-21.1.2	(SEQ ID NO:456)	49
YCGGDC	D2-21.2.2	(SEQ ID NO:457)	50
IVVVTa	D2-21.3.2	(SEQ ID NO:458)	51
LWWwLL	D2-21.1.3	(SEQ ID NO:465)	52
CGGDCY	D2-21.2.3	(SEQ ID NO:466)	53
VVVTAI	D2-21.3.3	(SEQ ID NO:467)	54
WWwLLF	D2-21.1.4	(SEQ ID NO:474)	55
GGDCYS	D2-21.2.4	(SEQ ID NO:475)	56
VLRFLF	D3-3.1.1	(SEQ ID NO:489)	57
YYDFWS	D3-3.2.1	(SEQ ID NO:490)	58
ITIFGV	D3-3.3.1	(SEQ ID NO:491)	59
LRFLEW	D3-3.1.2	(SEQ ID NO:498)	60

ES 2 528 963 T3

YDFWSG	D3-3.2.2	(SEQ ID NO:499)	61
TIFGVV	D3-3.3.2	(SEQ ID NO:500)	62
RFLEWL	D3-3.1.3	(SEQ ID NO:507)	63
DFWSGY	D3-3.2.3	(SEQ ID NO:508)	64
IFGVVI	D3-3.3.3	(SEQ ID NO:509)	65
FLEWLL	D3-3.1.4	(SEQ ID NO:516)	66
FWSGYY	D3-3.2.4	(SEQ ID NO:517)	67
FGVVII	D3-3.3.4	(SEQ ID NO:518)	68
LEWLLY	D3-3.1.5	(SEQ ID NO:525)	69
WSGYYT	D3-3.2.5	(SEQ ID NO:526)	70
VLRYFD	D3-9.1.1	(SEQ ID NO:539)	71
YYDILT	D3-9.2.1	(SEQ ID NO:540)	72
ITIFyL	D3-9.3.1	(SEQ ID NO:541)	73
LRYFDW	D3-9.1.2	(SEQ ID NO:548)	74
YDILTG	D3-9.2.2	(SEQ ID NO:549)	75
TIFyLV	D3-9.3.2	(SEQ ID NO:550)	76
RYFDWL	D3-9.1.3	(SEQ ID NO:557)	77
DILTGy	D3-9.2.3	(SEQ ID NO:558)	78
IFyLVI	D3-9.3.3	(SEQ ID NO:559)	79
YFDWLL	D3-9.1.4	(SEQ ID NO:566)	80
ILTGYY	D3-9.2.4	(SEQ ID NO:567)	81
FyLVII	D3-9.3.4	(SEQ ID NO:568)	82
FDWLLy	D3-9.1.5	(SEQ ID NO:575)	83
LTGYYN	D3-9.2.5	(SEQ ID NO:576)	84
VLLWFG	D3-10.1.1	(SEQ ID NO:590)	85
YYYGSG	D3-10.2.1	(SEQ ID NO:591)	86
ITMVRG	D3-10.3.1	(SEQ ID NO:592)	87
LLWFGE	D3-10.1.2	(SEQ ID NO:599)	88
YYGSGS	D3-10.2.2	(SEQ ID NO:600)	89
TMVRGV	D3-10.3.2	(SEQ ID NO:601)	90
LWFGEL	D3-10.1.3	(SEQ ID NO:608)	91
YGSGSY	D3-10.2.3	(SEQ ID NO:609)	92
MVRGVI	D3-10.3.3	(SEQ ID NO:610)	93
WFGELL	D3-10.1.4	(SEQ ID NO:617)	94
GSGSYY	D3-10.2.4	(SEQ ID NO:618)	95
VRGVII	D3-10.3.4	(SEQ ID NO:619)	96

ES 2 528 963 T3

FGELLy	D3-10.1.5	(SEQ ID NO:624)	97
SGSYYN	D3-10.2.5	(SEQ ID NO:625)	98
VLwLRL	D3-16.1.1	(SEQ ID NO:638)	99
YYDYVW	D3-16.2.1	(SEQ ID NO:639)	100
IMITFG	D3-16.3.1	(SEQ ID NO:640)	101
LwLRLG	D3-16.1.2	(SEQ ID NO:647)	102
YDYVWG	D3-16.2.2	(SEQ ID NO:648)	103
MITFGG	D3-16.3.2	(SEQ ID NO:649)	104
wLRLGE	D3-16.1.3	(SEQ ID NO:656)	105
DYVWGS	D3-16.2.3	(SEQ ID NO:657)	106
ITFGGV	D3-16.3.3	(SEQ ID NO:658)	107
LRLGEL	D3-16.1.4	(SEQ ID NO:665)	108
YVWGSY	D3-16.2.4	(SEQ ID NO:666)	109
TFGGVI	D3-16.3.4	(SEQ ID NO:667)	110
RLGELS	D3-16.1.5	(SEQ ID NO:674)	111
VWGSYR	D3-16.2.5	(SEQ ID NO:675)	112
FGGVIV	D3-16.3.5	(SEQ ID NO:676)	113
LGELSL	D3-16.1.6	(SEQ ID NO:683)	114
WGSYRY	D3-16.2.6	(SEQ ID NO:684)	115
GGVIVI	D3-16.3.6	(SEQ ID NO:685)	116
GELSLY	D3-16.1.7	(SEQ ID NO:692)	117
GSYRYT	D3-16.2.7	(SEQ ID NO:693)	118
VLLwyy	D3-22.1.1	(SEQ ID NO:707)	119
YYYDSS	D3-22.2.1	(SEQ ID NO:708)	120
ITMIVV	D3-22.3.1	(SEQ ID NO:709)	121
LLwyyW	D3-22.1.2	(SEQ ID NO:716)	122
YYDSSG	D3-22.2.2	(SEQ ID NO:717)	123
TMIVVV	D3-22.3.2	(SEQ ID NO:718)	124
LwyyWL	D3-22.1.3	(SEQ ID NO:725)	125
YDSSGY	D3-22.2.3	(SEQ ID NO:726)	126
MIVVVI	D3-22.3.3	(SEQ ID NO:727)	127
wyyWLL	D3-22.1.4	(SEQ ID NO:733)	128
DSSGYY	D3-22.2.4	(SEQ ID NO:734)	129
IVVVIT	D3-22.3.4	(SEQ ID NO:735)	130
yyWLLL	D3-22.1.5	(SEQ ID NO:742)	131
SSGYYY	D3-22.2.5	(SEQ ID NO:743)	132

wLRWyL	D4-23.1.1	(SEQ ID NO:769)	133
DYGGNS	D4-23.2.1	(SEQ ID NO:770)	134
VDTAMV	D5-5.1.1	(SEQ ID NO:784)	135
WIQLWL	D5-5.2.1	(SEQ ID NO:785)	136
GYSYGY	D5-5.3.1	(SEQ ID NO:786)	137
VDIVAT	D5-12.1.1	(SEQ ID NO:802)	138
WlyWLR	D5-12.2.1	(SEQ ID NO:803)	139
GYSGYD	D5-12.3.1	(SEQ ID NO:804)	140
DIVATI	D5-12.1.2	(SEQ ID NO:811)	141
lyWLRL	D5-12.2.2	(SEQ ID NO:812)	142
YSGYDY	D5-12.3.2	(SEQ ID NO:813)	143
VEMATI	D5-24.1.1	(SEQ ID NO:829)	144
YRWLQL	D5-24.2.1	(SEQ ID NO:830)	145
RDGYNY	D5-24.3.1	(SEQ ID NO:831)	146
EYSSSS	D6-6.1.1	(SEQ ID NO:847)	147
GYSSSW	D6-13.1.1	(SEQ ID NO:859)	148
GIAAAG	D6-13.2.1	(SEQ ID NO:860)	149
VyQQLV	D6-13.3.1	(SEQ ID NO:861)	150
YSSSWY	D6-13.1.2	(SEQ ID NO:867)	151
GYSSGW	D6-19.1.1	(SEQ ID NO:878)	152
GIAVAG	D6-19.2.1	(SEQ ID NO:879)	153
VyQWLV	D6-19.3.1	(SEQ ID NO:880)	154
YSSGWY	D6-19.1.2	(SEQ ID NO:887)	155

Ejemplo 3: CDR3 de longitud 6-20

La inserción de segmentos D en CDR3s de HC sintéticas puede conducir a una mayor estabilidad y a una menor inmunogenicidad. Se diseñan bibliotecas a nivel de aminoácidos uniendo un VH a un relleno opcional de cierta longitud que está unido a un segmento D, un segundo relleno opcional y un JH. Para bibliotecas de longitud seis u ocho, un JH de longitud completa puede seguir a VH y a un relleno corto. La Tabla 77 muestra la frecuencia de segmentos D en un muestreo de 1419 Abs seleccionados de FAB-310 o FAB-410 para la unión a una diana u otra. En la muestra, 1099 Abs no tuvieron segmento D detectable (es decir, menos de 70% de coincidencia). Cuando se usaron segmentos D, se prefieren los segmentos D D1-1.3, D1-26.3, D2-2.2, D2-8.2, D2-15.2, D2-21.2, D3-16.2, D3-22.2, D3-3.2, D3-9.1, D3-9.2, D3-10.2, D3-16.2, D4-4.2, D4-4.3, D4-11.2, D4-4.2, D4-17.2, D4-23.2, D5-5.3, D5-12.3, D5-18.3, D6-6.1, D6-6.2, D6-13.1, D6-13.2, D6-19.1, D6-19.2, y D7-27.1.

Una vez que se ha diseñado la secuencia de aminoácidos progenitora, se puede diversificar de varias maneras: PCR con tendencia a error, titubeo, y titubeo digital. La Tabla 14 muestra un número de hexámeros que se pueden derivar a partir de regiones D humanas. En una realización, se excluyen los hexámeros que contienen restos de cisteína. En una realización, se excluyen los fragmentos de regiones D que contienen paradas. En una realización, cualquier codón TAG encontrado en la región D se sustituye por un codón recogido del conjunto que comprende TCG, TTG, TGG, CAG, AAG, TAT, y GAG. En una realización, cualquier codón TAA encontrado en la región D se sustituye por un codón recogido del conjunto que comprende TCA, TTA, CAA, AAA, TAT, y GAA. En una realización, cualquier TGA de la región D se sustituye por un codón recogido del conjunto que comprende TGG, TCA, TTA, AGA, y GGA.

La Tabla 21 muestra secuencias de aminoácidos progenitoras ejemplares para CDR3s de 6 a 20 aminoácidos. Estas

5 secuencias progenitoras se pueden combinar con diversidad en CDR1 y CDR2 de HC para formar una biblioteca. Es probable que la utilidad mejore si las regiones de CDR3 se diversifican, por ejemplo, mediante titubeo, titubeo digital, o PCR con tendencia a error de las CDR3s. En la Tabla 21, la secuencia 6a comprende el extremo de VH de 3-23 fusionado a JH1 completo. La secuencia 6b contiene el extremo de 3-23 unido a un Y unido a D4-17 (RF 2) unido a la región FR4 de JH1. La secuencia 6c contiene el extremo de 3-23 seguido de D5-5 (RF 3) seguido de la parte FR4 de JH1. La secuencia 6d contiene el extremo de 3-23 unido a SY unido a JH4 completo. La Tabla 21 muestra el nivel de dopaje que sería el apropiado para el titubeo de la CDR3; igualmente se podrían usar otros niveles. Se podrían usar otras regiones D o fragmentos de regiones D. Se podrían usar otras secuencias de JH.

Tabla 21: Secuencias de aminoácidos progenitoras para CDR3s de HC de 6-20 AAs

Longitud	Secuencia progenitora	Nivel de dopaje	Comentario	SEQ ID NO:
6a	yycakAEYFQHWgqgltvtvss	70:10:10:10	JH1(completo)	226
6b	yycakYDYGDYwgggltvtvss	70:10:10:10	Y::D4-17(2)::FR4 de JH1	227
6c	yycakGYSGYwgggltvtvss	70:10:10:10	D5-5(3)::FR4 de JH1	228
6d	yycakSYFDYwgggltvtvss	70:10:10:10	SY::JH4(completo)	229
8a	yycakYYAEYFQHWgqgltvtvss	73:9:9:9	YY:JH1(completo)	230
8b	yycakYGYSSSWYwgggltvtvss	73:9:9:9	Y::D6-13(1)::FR4 de JH1	231
8c	yycakYGDYFDYwgggltvtvss	73:9:9:9	D4-17(2)[2-5]::JH4(completo)	232
10a	yycakYYDSSSGYYYwgggltvtvss	73:9:9:9	D3-22(2)::FR4 de JH1	233
10b	yycakGYcSSTScYTwgggltvtvss	73:9:9:9	D2-2(2)::FR4 de JH1	234
10c	yycakYYSSAEYFQHWgqgltvtvss	73:9:9:9	YYSS::JH1(completo)	235
10d	yycakGYSGYFDYwgggltvtvss	73:9:9:9	DS-5(3)::JH4(completo)	236
12a	yycakYYDSSSGYYYQHwgggltvtvss	85:5:5:5	D3-22(2)::QH::FR4 de JH1	237
12b	yycakGYcSSTScYQHWgqgltvtvss	85:5:5:5	D2-2(2)::QH:FR4 de JH1	238
12c	yycakYDGSYSAEYFQHWgqgltvtvss	85:5:5:5	YDGSYS::JH1(completo)	239
12d	yycakYDYVWGSYRYTwgggltvtvss	85:5:5:5	D3-16(2)::Fr de JH1	240
12e	yycakGYSGYWYFDLwgrgltvtvss	85:5:5:5	D5-5(3)::JH2(completo)	241
14a	yycakYYDSSSGYYYFQHWgqgltvtvss	73:9:9:9	D3-22(2)::YFQH::Fr de JH1	242
14b	yycakGYcSSTScYTYFQHWgqgltvtvss	73:9:9:9	D2-2(2)::YFQH:Fr de JH1	243
14c	yycakSYGYcSSTScYTYQHWgqgltvtvss	73:9:9:9	SY::D2-2(2)::QH::Fr de JH1	244
14d	yycakSYRYSGYSAEYFQHWgqgltvtvss	73:9:9:9	SYRYSGYS::JH1(completo)	245
14e	yycakAYcGGDcYSNWFDpwgggltvtvss	73:9:9:9	D2-21(2)::JH5(completo)	246
15a	yycakSDGYYYDSSSGYYYDYwgggltvtvss	73:9:9:9	SD::D3-22.2::JH4(101ff)	930

Longitud	Secuencia progenitora	Nivel de dopaje	Comentario	SEQ ID NO:
15b	yycakGGYCSGGSCYSFDYwgqgltvtss	73:9:9:9	GS::D2-15.2::JH4(100ff)	931
15c	yycakGGRGYSSGWYRAFDlwgqgtnvtss	73:9:9:9	GGR::D6-19.1::R::JH3(todo)	932
16a	yycakYYYDSSGYYYAEYFQHwgqgltvtss	73:9:9:9	D3-22(2)::JH1 (completo)	247
16b	yycakGYcSSTScYTAEYFQHwgqgltvtss	73:9:9:9	D2-2(2)::JH1 (completo)	248
16c	yycakSYDSYRSYGSAAEYFQHwgqgltvtss	73:9:9:9	SYDSYRSYGS::JH1(completo)	249
16d	yycakSYSYGYcSSTScYTQHwgqgltvtss	73:9:9:9	SYSY::D2-2(2)::QH::Fr JH1	250
17a	yycakSRPGYSSSWYYYGMDVwgqgltvtss	73:9:9:9	SRP::6-13.1::JH6(-1Y)	933
18a	yycakGYCSGGSCYSYYYGMDVwgqgltvt	73:9:9:9	2-15.2::1H6(-1Y)	221
18b	yycakDGYCSGGSCYSYYYGMDVwgqgltvt	73:9:9:9	D::2-15.2::JH6(-2Ys)	222
19a	yycakDGYYYDSSGYYYRGYFYFDYwgqgltvt	73:9:9:9	D::D3-22.2::RGY::JH4(todo)	223
20a	yycakYSSYYYDSSGYYYAEYFQHwgqgltvt	73:9:9:9	YSSY::D3-22(2)::JH 1 (completo)	251
20b	yycakSYSGYcSSTScYTAEYFQHwgqgltvt	73:9:9:9	SYYS::D2-2(2)::JH 1 (completo)	252
20c	yycakSGYcSSTScYTYSAEYFQHwgqgltvt	73:9:9:9	S::D2-2(2)::YYS::JH1(completo)	253
20d	yycakYYYDYVWGSYRYTSNWFDPwgqgltvt	73:9:9:9	Y::D3-16(2)::S::JH5(completo)	254
20e	yycakYYYDYVWGSYRYTSSYFDYwgqgltvt	73:9:9:9	Y::D3-16(2)::SS::JH4(completo)	255

Tabla 22: Casete de presentación de HC

La secuencia de aminoácidos mostrada en la Tabla 22 es SEQ ID NO:892.

La secuencia de ADN mostrada en la Tabla 22 es SEQ ID NO:893.

```

!      señal para VH-CH1-IIIstump
!      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
!      M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L
!  946  atg aaa tac cta ttg cct acg gca gcc gct gga ttg tta tta ctc
!
!      16 17 18 19 20 21 22
!      A  A  Q  P  A  M  A
!  991  gcG GCC cag ccG GCC atg gcc
!      SfiI.....
!      NgoMI... (1/2)
!      NcoI....
!
!  VH
!
!      FR1 (DP47/V3-23)-----
!      1  2  3  4  5  6  7  8
!      E  V  Q  L  L  E  S  G
!  1012  gaa|ggt|CAA|TTG|tta|gag|tct|ggt|
!      | MfeI |
!
!      -----FR1-----
!      9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
!      G  G  L  V  Q  P  G  G  S  L  R  L  S  C  A
!  1036  |ggc|ggg|ctt|ggt|cag|cct|ggt|ggt|tct|tta|cgt|ctt|tct|tgc|gct|
!
!      -----FR1----->|...CDR1.....|-----FR2-----
!      24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
!      A  S  G  F  T  F  S  S  Y  A  M  S  W  V  R
!  1081  |gct|TCC|GGA|ttc|act|ttc|tct|tCG|TAC|Gct|atg|tct|tgg|ggt|cgC|
!      | BspEI |          | BsiWI |          |BstXI.
!
!      -----FR2----->|...CDR2.....
!      39 40 41 42 43 44 45 46 47 48 49 50 51 52 52a
!      Q  A  P  G  K  G  L  E  W  V  S  A  I  S  G
!  1126  |CAa|gct|ccT|GGT|aaa|ggt|ttg|gag|tgg|ggt|tct|gct|atc|tct|ggt|
!      ...BstXI          |
!
!      ....CDR2.....|-----FR3-----
!      53 54 55 56 57 58 59 60 61 62 63 64 65 66 67
!      S  G  G  S  T  Y  Y  A  D  S  V  K  G  R  F
!  1171  |tct|ggt|ggc|agt|act|tac|tat|gct|gac|tcc|ggt|aaa|ggt|cgc|ttc|
!
!      -----FR3-----
!      68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
!      T  I  S  R  D  N  S  K  N  T  L  Y  L  Q  M
!  1216  |act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|
!      | XbaI |
!
!      -----FR3----->|
!  82a 82b 82c 83 84 85 86 87 88 89 90 91 92 93 94

```

ES 2 528 963 T3

La secuencia de aminoácidos mostrada en la Tabla 22 es SEQ ID NO:892.

La secuencia de ADN mostrada en la Tabla 22 es SEQ ID NO:893.

```

!      N   S   L   R   A   E   D   T   A   V   Y   Y   C   A   K
1261 |aac|agC|TTA|AGg|gct|gag|gac|aCT|GCA|Gtc|tac|tat|tgc|gct|aaa|
!      |AflIII |          | PstI | (2/2)
!
!      .....CDR3.....|----FR4-----
!      95  96  97  98  98a 98b 98c  99  100 101 102 103 104 105 106
!      D   Y   E   G   T   G   Y   A   F   D   I   W   G   Q   G
1306 |gac|tat|gaa|ggg|act|ggg|tat|gct|ttc|gaC|ATA|TGg|ggg|caa|ggg|
!      | NdeI |
!
!      -----FR4----->|
!      107 108 109 110 111 112 113
!      T   M   V   T   V   S   S
1351 |act|atG|GTC|ACC|gtc|tct|agt
!      | BstEII | c tcg ag = XhoI.
!
! CH1
!
!      A   S   T   K   G   P   S   V   F   P   L   A   P   S   S
1372 |gcc|tcc|acc|aag|ggc|cca|tcg|gtc|ttc|ccG|CTA|GCa|ccc|tcc|tcc
!      | NheI....
!
!      151 152 153 154 155 156 157 158 159 160 161 162 163 164 165
!      K   S   T   S   G   G   T   A   A   L   G   C   L   V   K
1417 |aag|agc|acc|tct|ggg|ggc|aca|gcg|gcc|ctg|ggc|tgc|ctg|gtc|aag
!
!      166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
!      D   Y   F   P   E   P   V   T   V   S   W   N   S   G   A
1462 |gac|tac|ttc|ccc|gaa|ccg|gtg|acg|gtg|tcg|tgg|aac|tca|ggc|gcc
!
!      181 182 183 184 185 186 187 188 189 190 191 192 193 194 195
!      L   T   S   G   V   H   T   F   P   A   V   L   Q   S   S
1507 |ctg|acc|agc|ggc|gtc|cac|acc|ttc|ccg|gct|gtc|cta|cag|tcc|tca
!
!      196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
!      G   L   Y   S   L   S   S   V   V   T   V   P   S   S   S
1552 |gga|ctc|tac|tcc|ctc|agc|agc|gta|gtg|acc|gtg|ccc|tCC|Agc|agc
!      | BstXI.....
!
!      211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
!      L   G   T   Q   T   Y   I   C   N   V   N   H   K   P   S
1597 |tTG|Ggc|acc|cag|acc|tac|atc|tgc|aac|gtg|aat|cac|aag|ccc|agc
!      | BstXI.....
!
!      226 227 228 229 230 231 232 233 234 235 236 237 238
!      N   T   K   V   D   K   K   V   E   P   K   S   C
1642 |aac|acc|aag|gtg|gac|aaG|AAA|GTT|GAG|CCC|AAA|TCT|TGT
!
!      139 140 141 His tag..... cMyc tag.....
!      A   A   A   H   H   H   H   H   H   G   A   A   E   Q   K   L   I
1681 |GCG|GCC|GCa|cat|cat|cat|cac|cat|cac|ggg|gcc|gca|gaa|caa|aaa|ctc|atc
!      | NotI.....
!      | EagI....
!
!      .....
!      S   E   E   D   L   N   G   A   A   E   A   S   S   A   S   N   A   S
1732 |tca|gaa|gag|gat|ctg|aat|ggg|GCC|gca|gaG|Gct|agt|tct|gct|agt|aAC|GCG|Tct

```

La secuencia de aminoácidos mostrada en la Tabla 22 es SEQ ID NO:892.

La secuencia de ADN mostrada en la Tabla 22 es SEQ ID NO:893.

```

!                                     BglI..... (3/4)                               MluI....
!
! Dominio 3 (IIIstump)-----
!   S   G   D   F   D   Y   E   K   M   A   N   A   N   K   G   A
! 1786 tcc ggt gat ttt gat tat gaa aag atg gca aac gct aat aag ggg gct
!
!   M   T   E   N   A   D   E   N   A   L   Q   S   D   A   K   G
! 1834 atg acc gaa aat gcc gat gaa aac gcg cta cag tct gac gct aaa ggc
!
!   K   L   D   S   V   A   T   D   Y   G   A   A   I   D   G   F
! 1882 aaa ctt gat tct gtc gct act gat tac ggt gct gct atc gat ggt ttc
!
!   I   G   D   V   S   G   L   A   N   G   N   G   A   T   G   D
! 1930 att ggt gac gtt tcc ggc ctt gct aat ggt aat ggt gct act ggt gat
!
!   F   A   G   S   N   S   Q   M   A   Q   V   G   D   G   D   N
! 1978 ttt gct ggc tct aat tcc caa atg gct caa gtc ggt gac ggt gat aat
!
!   S   P   L   M   N   N   F   R   Q   Y   L   P   S   L   P   Q
! 2026 tca cct tta atg aat aat ttc cgt caa tat tta cct tcc ctc cct caa
!
!   S   V   E   C   R   P   F   V   F   G   A   G   K   P   Y   E
! 2074 tcg gtt gaa tgt cgc cct ttt gtc ttt ggc gct ggt aaa cca tat gaa
!
!   F   S   I   D   C   D   K   I   N   L   F   R
! 2122 ttt tct att gat tgt gac aaa ata aac tta ttc cgt
!                                     Extremo dominio 3
!
!   G   V   F   A   F   L   L   Y   V   A   T   F   M   Y   V   F140
! 2158 ggt gtc ttt gcg ttt ctt tta tat gtt gcc acc ttt atg tat gta ttt
! comienzo segmento transmembránico
!
!   S   T   F   A   N   I   L
! 2206 tct acg ttt gct aac ata ctg
!
!   R   N   K   E   S   (SEQ ID NO:892)
! 2227 cgt aat aag gag tct TAA   tga aAC GCG Tga tga GAATTC (SEQ ID NO:893)
! Anclaje intracelular                               MluI....                               EcoRI.

```

Tabla 25: La secuencia de ADN de DY3F85LC que contiene una cadena ligera kappa O12 de línea germinal de muestra. Las secuencias del anticuerpo mostradas son de la forma de anticuerpo real, pero no se han identificado como unidas a un antígeno particular.

En cada línea, todo lo que está después de un signo de admiración (!) es un comentario.

El ADN de DY3F85LC es SEQ ID NO:27

```

!-----
! 1      AATGCTACTA CTATTAGTAG AATIGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT
! 61     ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT
! 121    CGTTCGCAGA ATTGGGAATC AACIGTTATA TGGAATGAAA CTTCCAGACA CCGTACTTTA
! 181    GTTGCAATAT TAAAACATGT TGAGCTACAG CATTATATTC AGCAATTAAAG CTCTAAGCCA
! 241    TCCGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG
! 301    TTGGAGTTTG CTTCCGGTCT GGTTGCTTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG
! 361    TCTTTCGGGC TTCCTCTTAA TCTTTTGTAT GCAATCCGCT TTGCTTCTGA CTATAATAGT
! 421    CAGGGTAAAG ACCTGATTTT TGATTATGTT TCATTCTCGT TTTCTGAACG GTTTAAAGCA

```

En cada línea, todo lo que está después de un signo de admiración (!) es un comentario.

El ADN de DY3F85LC es SEQ ID NO:27

```

481      TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT
541      AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT
601      GGTTTTTATC GTCGCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT
661      AATTCCTTTT GCGGTTATGT ATCTGCATTA GTTGAATGTG GTATTCTTAA ATCTCAACTG
721      ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT
781      TCTTCCCAAC GTCTTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA
841      CAATGATTAA AGITGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT
901      CTCGTAGGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG
961      AATATCCGGT TCITGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC
1021     TGTACACCGT TCATCTGTCC TCTTTCAAAAG TTGGTCAGTT CGGTTCCTT ATGATTGACC
1081     GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTCC CGGATTTTCA CACAATTTAT
1141     CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTCGCGC TTGGTATAAT CGCTGGGGGT
1201     CAAAGATGAG TGTTTTAGTG TATTCTTTTG CCTCTTTTCT TTTAGGTTGG TGCCCTTCGTA
1261     GTGGCATTAC GTATTTTACC CGTTTAAATG AAACCTCCTC ATGAAAAAGT CTTTAGTCCT
1321     CAAAGCCTCT GTAGCCGTTG CTACCTCTCG TCCGATGCTG TCTTTCGCTG CTGAGGGTGA
1381     CGATCCCGCA AAAGCGGCCT TTAACCTCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA
1441     TGCCTGGGCG ATGGTTGTTG TCATTGTCTG CGCAACTATC GGTATCAAGC TGTTAAGAA
1501     ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT
1561     TTTTGGAGAT TTTCAACGTG AAAAAATTAT TATTCGCAAT TCCTTTAGTT GTTCCTTTCT
1621     ATTCTCACTC CGCTGAAACT GTTGAAAGTT GTTTAGCAAA ATCCCATACA GAAAATTCAT
1681     TTAATAACGT CTGGAAAGAC GACAAAACCT TAGATCGTTA CGCTAACTAT GAGGGCTGTC
1741     TGTGGAATGC TACAGGCGTT GTAGTTTGTA CTGGTGACGA AACTCAGTGT TACGGTACAT
1801     GGGTTCTTAT TGGGCTTGCT ATCCCTGAAA ATGAGGGTGG TGGCTCTGAG GGTGGCGGTT
1861     CTGAGGGTGG CGGTTCGTAG GGTGGCGGTA CTAACCTCC TGAGTACGGT GATACACCTA
1921     TTCCGGGCTA TACTTATATC AACCTCTCTG ACGGCACCTA TCCGCTGGT ACTGAGCAAA
1981     ACCCCGCTAA TCCTAATCCT TCCTTTGAGG AGTCTCAGCC TCTTAATACT TTCATGTTTC
2041     AGAATAATAG GTTCCGAAAT AGGCAGGGGG CATTAACTGT TTATACGGGC ACTGTTACTC
2101     AAGGCACCTG CCCCGTTAAA ACTTATTACC AGTACACTCC TGTATCATCA AAAGCCATGT
2161     ATGACGCTTA CTGGAACGGT AAATTCAGAG ACTGCGCTTT CCAITCTGGC TTTAATGAGG
2221     ATTTATTGTG TTGTGAATAT CAAGGCCAAT CGTCTGACCT GCCTCAACCT CCTGTCAATG
2281     CTGGCGGCGG CTCTGGTGGT GGTCTGGTG GCGGCTCTGA GGGTGGTGGC TCTGAGGGTG
2341     GCGGTTCGTA GGGTGGCGGC TCTGAGGGAG GCGGTTCGGG TGGTGGCTCT GGTTCGGGTG
2401     ATTTTGATTA TGAAAGATG GCAAACGCTA ATAAGGGGGC TATGACCGAA AATGCCGATG
2461     AAAACGCGCT ACAGTCTGAC GTAAAGGCA AACTTGATTG TGTCGCTAT GTTACGGTG
2521     CTGCTATCGA TGGTTTCATT GGTGACGTTT CCGGCCCTGC TAATGGTAAT GGTGCTACTG
2581     GTGATTTTGC TGGCTCTAAT TCCCAAATGG CTCAAGTCCG TGACGGTGAT AATTCACCTT
2641     TAATGAATAA TTTCCGTCAA TATTTACCTT CCCTCCCTCA ATCGGTTGAA TGTCCGCCCT
2701     TTGTCTTTGG CGCTGGTAAA CCATATGAAT TTTCTATTGA TTGTGACAAA ATAAACTTAT
2761     TCCGTGGTGT CTTTGGCTTT CTTTATATG TTGCCACCTT TATGTATGTA TTTTCTACGT
2821     TTGCTAACAT ACTGCGTAAT AAGGAGTCTT AATCATGCCA GTTCTTTTGG GTATTCCGTT
2881     ATTATTGCGT TTCCTCGGTT TCCTTCTGGT AACTTTGTTC GGCTATCTGC TTACTTTTCT
2941     TAAAAAGGGC TTCGGTAAGA TAGCTATTGC TATTTTCATTG TTTCTTGCTC TTATTATTGG
3001     GCTTAACTCA ATTCTTGTGG GTTATCTCTC TGATATTAGC GCTCAATTAC CCTCTGACTT
3061     TGTTACGGGT GTTCAGTTAA TTCTCCCGTC TAATGCGCTT CCCTGTTTTT ATGTTATTCT
3121     CTCTGTAAAG GCTGCTATTT TCATTTTGA CGTTAAACAA AAAATCGTTT CTTATTTGGA
3181     TTGGGATAAA TAATATGGCT GTTTATTTTG TAACTGGCAA ATTAGGCTCT GGAAGACGCG
3241     TCGTTAGCGT TGGTAAGATT CAGGATAAAA TTGTAGCTGG GTGCAAAATA GCAACTAATC
3301     TTGATTTAAG GCTTCAAAAC CTCCCGCAAG TCGGGAGGTT CGCTAAAACG CCTCGCGTTC
3361     TTAGAATACC GGATAAGCCT TCTATATCTG ATTTGCTTGC TATTGGGCGC GGTAAATGAT
3421     CCTACGATGA AAATAAAAAC GGCTTGTCTG TTCTCGATGA GTGCGGTACT TGGTTTAATA
3481     CCCGTTCTTG GAATGATAAG GAAAGACAGC CGATTATTGA TTGGTTTCTA CATGCTCGTA
3541     AATTAGGATG GGATATTATT TTTCTTGTTC AGGACTTATC TATTGTTGAT AAACAGGCGC
3601     GTTCTGCATT AGCTGAACAT GTTGTTTTAT GTCTGCTCTT GGACAGAATT ACTTTACCTT
3661     TTGTCGGTAC TTTATATTCT CTTATTACTG GCTCGAAAAT GCCTCTGCC TAAATTACATG
3721     TTGGCGTTGT TAAATATGGC GATTCTCAAT TAAGCCCTAC TGTGAGCGT TGGCTTTATA
3781     CTGGTAAGAA TTTGTATAAC GCATATGATA CTAAACAGGC TTTTCTAGT AATTATGATT
3841     CCGGTGTTTA TTCTTATTTA ACGCCTTATT TATCACACGG TCGGTATTTT AAACCATTA

```

En cada línea, todo lo que está después de un signo de admiración (!) es un comentario.

El ADN de DY3F85LC es SEQ ID NO:27

```

3901      ATTTAGGTCA GAAGATGAAA TTAATAAAAA TATATTTGAA AAAGTTTTCT CGCGTTCCTT
3961      GTCTTGCGAT TGGATTGCA TCAGCATTTA CATATAGTTA TATAACCCAA CCTAAGCCGG
4021      AGGTTAAAAA GGTAGTCTCT CAGACCTATG ATTTTGATAA ATTCACATT GACTCTTCTC
4081      AGCGTCTTAA TCTAAGCTAT CGCTATGTTT TCAAGGATTC TAAGGGAAAA TTAATTAATA
4141      GCGACGATTT ACAGAAGCAA GGTATTTCAC TCACATATAT TGATTTATGT ACTGTTTCCA
4201      TTAATAAAGG TAATTCAAAT GAAATGTGTA AATGTAATTA ATTTTGTTTT CTGTATGTTT
4261      GTTTCATCAT CTCTTTTTGC TCAGGTAATT GAAATGAATA ATTCGCTCT GCGCGATTTT
4321      GTAACCTGGT ATTCAAAGCA ATCAGGCGAA TCCGTTATTG TTTCTCCCGA TGTAAGAAGG
4381      ACTGTTACTG TATATTCATC TGACGTTAAA CCTGAAAAATC TACGCAATTT CTTTATTTCT
4441      GTTTTACGTG CAAATAATTT TGATATGGTA GGTCTAACC CTTCATAAT TCAGAAGTAT
4501      AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT
4561      GATAATTCGG CTCCTTCTGG TGGTTTCTTT GTTCCGCAA ATGATAATGT TACTCAAAC
4621      TTTAAAAATTA ATAACGTTCG GGCAAGGAT TTAATACGAG TTGTCGAATT GTTTGTAAAG
4681      TCTAATACTT CTAAATCCTC AAATGTATTA TCTATTGACG GCTCTAATCT ATTAGTTTGA
4741      AGTGCTCCTA AAGATATTTT AGATAACCTT CCTCAATTCC TTTCAACTGT TGATTTGCCA
4801      ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTTT AGCAAGGTGA TGCTTTAGAT
4861      TTTTCATTTG CTGCTGGCTC TCAGCGTGGC ACTGTTGCAG GCGGTGTTAA TACTGACCGC
4921      CTACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTCCGTA TTTTAAATGG CGATGTTTAA
4981      GGGCTATCAG TTCGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT
5041      ATTCTTACGC TTTTCAGGTA GAAGGGTTCT ATCTCTGTTG GCCAGAAATG CCCTTTTATT
5101      ACTGGTCTGT TGACTGGTGA ATCTGCCAAT GTAAATAATC CATTTACAGC GATTGACGCT
5161      CAAAATGTAG GTATTTCCAT GAGCGTTTTT CCTGTTGCAA TGGCTGGCGG TAATATTGTT
5221      CTGGATATTA CCAGCAAGGC CGATAGTTTG AGTTCTTCTA CTCAGGCAAG TGATGTTATT
5281      ACTAATCAAA GAAGTATTGC TACAACGGTT AATTGCGTG ATGGACAGC TCTTTTACTC
5341      GGTGGCCTCA CTGATTATAA AAACACTTCT CAGGATTCTG GCGTACCGTT CCGTCTTAAA
5401      ATCCCTTTAA TCGGCCCTCT GTTTAGCTCC CGCTCTGATT CTAAACGAGG AAGCAGGTTA
5461      TACGTGCTCG TCAAAGCAAC CATAGTACGC GCCCTGTAGC GCGCGATTAA GCGCGGCGGG
5521      TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT
5581      CGCTTCTTTC CCTTCCCTTC TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG
5641      GGGGCTCCCT TTAGGGTTTC GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAAGTTGA
5701      TTTGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTGAC
5761      GTTGGAGTCC ACGTTCCTTA ATAGTGGACT CTTGTTCCAA ACTGGAACAA CACTCAACCC
5821      TATCTCGGGC TATTTCTTTG ATTTATAAGG GATTTTGCCG ATTTCCGAAC CACCATCAAA
5881      CAGGATTTTC GCCTGCTGGG GCAAACAGC GTGGACCGCT TGCTGCAACT CTCTCAGGCG
5941      CAGGCGGTGA AGGGCAATCA GCTGTTGCC GTCTCACTGG TGAAGAAGAA AACCACCTG
6001      GATCCAAGCT TGCAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT
6061      TTTTCTAAAT ACATTCAAA ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC
6121      AATAATATTG AAAAAGGAAG AGTATGAGTA TTCAACATT CCGTGTCCGC CTATTCCCT
6181      TTTTGCGGC ATTTTGCTT CCTGTTTTG CTCACCCAGA AACGCTGTG AAAGTAAAG
6241      ATGCTGAAGA TCAGTTGGG GCACTAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA
6301      AGATCCTTGA GAGTTTTTCG CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTTAAAGTTC
6361      TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTGCGCGCA
6421      TACACTATTC TCAGAAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG
6481      ATGGCATGAC AGTAAGAGAA TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG
6541      CCAACTTACT TCTGACAACG ATCGGAGGAC CGAAGGAGCT AACCGTTTTT TTGCACAACA
6601      TGGGGGATCA TGTAACTCGC CTTGATCGTT GGGAAACCGA GCTGAATGAA GCCATACCAA
6661      ACGACGAGCG TGACACCACG ATGCTGTAG CAATGGCAAC AACGTTGCGC AAACATTATA
6721      CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAATTAAT AGACTGGATG GAGGCGGATA
6781      AAGTGCAGG ACCACTTCTG CGCTCGGCC TTCCGGCTGG CTGTTTTATT GCTGATAAAT
6841      CTGGAGCCGG TGAGCGTGGG TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC
6901      CCTCCGTAT CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA
6961      GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAACGTGCA GACCAAGTTT
7021      ACTCATATAT ACTTTAGATT GATTTAAAC TTCAATTTTA ATTTAAAGG ATCTAGGTGA
7081      AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG TTCCAGTGTA
7141      CGTAAGACCC CCAAGCTTGT CGACTGAATG GCGAATGGCG CTTTGCCTGG TTTCCGGCAC
7201      CAGAAGCGGT GCCGGAAGC TGGCTGGAGT GCGATCTTCC TGACGCTCGA GCGCAACGCA
!
xhoI...

7261      ATTAATGTGA GTTAGCTCAC TCATTAGGCA CCCAGGCTT TACACTTTAT GCTTCCGGCT
7321      CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
7381      GATTACGCCA AGCTTTGGAG CCTTTTTTTT GGAGATTTTC AAC

```


Tabla 30: Secuencia de ADN de DY3FHC87 (SEQ ID NO:894)

```

1  aatgctacta ctattagtag aattgatgcc accttttcag ctcgcgcccc aaatgaaaat
61  atagctaAAC aggttattga ccatttgcgga aatgtatcta atgggtcaaac taaatctact
121 cgttcgcaga attgggaatc aactgttata tggaaatgaaa cttccagaca ccgtacttta
181 gttgcatatt taaaacatgt tgagctacag cattatattc agcaattaaag ctctaagcca
241 tccgcaaaaa tgacctctta tcaaaaggag caattaaagg tactctctaa tctgacctg
301 ttggagtttg cttccggtct ggttcgcttt gaagctcgaa ttaaaacgcg atatttgaag
361 tctttcgggc ttccctctta tctttttgat gcaatccgct ttgcttctga ctataatagt
421 cagggtaaa acctgatttt tgatttatgg tcatctcgt tttctgaact gtttaaagca
481 tttagggggg attcaatgaa tatttatgac gattccgcag tattggacgc tatccagctc
541 aaacatttta ctattacccc ctctggcaaa acttcttttg caaaagcctc tcgctatttt
601 ggttttttatc gtcgtctggt aaacgagggg tatgatagt ttgctcttac tatgcctcgt
661 aatttcctttt ggcgttatgt atctgcatta gttgaatgtg gtattcctaa atctcaactg
721 atgaatcttt ctacctgtaa taatgttgtt ccgttagttc gttttattaa cgtagatttt
781 tcttcccaac gtcctgactg gtataatgag ccagttctta aaatcgcata aggtaatca
841 caatgattaa agttgaaatt aaacctctc aagcccaatt tactactcgt tctgggtgtt
901 ctgctcaggg caagccttat tcaactgaat agcagctttg ttacgttgat ttgggtaatg
961 aatatccggt tcttgtcaag attactcttg atgaaggcca gccagcctat gcgcctggtc
1021 tgtacaccgt tcatctgtcc tcttcaaaag ttggtcagtt cgggtccctt atgattgacc
1081 gtctgcgctt cggttcggct aagtaacatg gagcaggtcg cggatttcga cacaatttat
1141 cagggcgatga tacaatctc cgttgtactt tgtttcgcgc ttggtataat cgtcgggggt
1201 caaagatgag tggttttagt tattcttttg cctctttcgt tttaggttg tgccctcgta
1261 gtggcattac gtattttacc cgtttaatgg aaacttctc atgaaaaagt ctttagtctc
1321 caaagcctct gtagccgttg ctaccctcgt tccgatgctg tctttcgctg ctgaggggtga
1381 cgatcccgca aaagcggcct ttaactcctt gcaagcctca gcgacccaat atatcggtta
1441 tgcgtgggag atggttgttg tcaattgtcg cgcaactatc ggatccaagc tgtttaagaa
1501 attcacctcg aaagcaagct gataaaccca tacaattaaa ggctcctttt ggagcctttt
1561 tttttggaga ttttcaacgt gaaaaaatta ttattcgcaa ttctcttagt tgttcccttc
1621 tattctcact ccgctgaaac tgttgaagt tgtttagcaa aatcccatac agaaaattca
1681 tttactaacg tctggaaga cgacaaaact ttagatcgtt acgctaacta tgagggtgtg
1741 ctgtggaatg ctacaggcgt tgtagtgtgt actggtgacg aaactcagtg ttacgggtaca
1801 tgggttccta ttgggcttgc tatccctgaa aatgaggggt gtggctctga ggggtggcgtg
1861 tctgaggggt gcggttctga ggggtggcgt actaaacctc ctgagtacgg tgatacacct
1921 attccgggct atacttatat caacctctc gacggcactt atccgcctgg tactgagcaa
1981 aaccccgcta atcctaatac ttctcttgag gagtctcagc ctcttaatac tttcatgttt
2041 cagaataata ggttccgaaa taggcagggg gcattaaact tttatacggg cactgttact
2101 caaggcactg accccgttaa aacttattac cagtacactc ctgtatcatc aaaagccatg
2161 tatgacgctt actggaacgg taaattcaga gactgcgctt tccattctgg ctttaatgag
2221 gatatttttg tttgtgaata tcaaggccaa tegtctgacc tgcctcaacc tctgtcaat
2281 gctggcgcg gctctggttg tggttctggt ggcggtctg aggggtggtg ctctgagggg
2341 ggcggttctg aggggtggcg ctctgagggg ggcggttccg gtggtggctc tggttccggg
2401 gattttgatt atgaaaagat ggcaaacgct aat'aaggggg ctatgaccga aaatgccgat
2461 gaaaacgcgc tacagtctga cgctaaaggc aaacttgatt ctgtcgctac tgattacggg
2521 gctgctatcg atggtttcat tggtgacgtt tccggccttg ctaatggtaa tgggtgctact
2581 ggtgattttg ctggctctaa ttcccaaatg gctcaagtgc gtgacgggtg taattcacct
2641 ttaatgaata atttccgtca atatttacct tccctccctc aatcgggttg atgtcgccct
2701 tttgtctttg gcgctggtaa accatatgaa ttttctattg attgtgacaa aataaactta
2761 ttccgtgggt tctttgcgtt tcttttatat gttgccacct ttatgtatgt attttctacg
2821 tttgctaaca tactgcgtaa taaggagtct taatcatgcc agttcttttg ggtattccgt
2881 tattattgcy tttctcgggt ttctctctgg taactttggt cggctatctg ctacttttc
2941 ttaaaaaggg cttcggtaag atagctattg ctatttcatt gtttcttgct cttattattg

```

```

3001 ggcttaactc aattcttctg ggttatctct ctgatattag cgctcaatta cccctctgact
3061 ttgttcaggg tgttcagtta attctcccgt ctaatgcgct tccctgtttt tatgttattc
3121 tctctgtaaa ggctgctatt ttcatttttg acgttaaaca aaaaatcggt tcttattttg
3181 attgggataa ataatatggc tgtttatttt gtaactggca aattaggctc tggaaaagacg
3241 ctcgtttagc ttggtaagat tcaggataaa attgtagctg ggtgcaaaat agcaactaat
3301 cttgatttaa ggcttcaaaa cctcccgcga gtcgggaggt tcgctaaaaa gcctcgcgtt
3361 cttagaatac cggataagcc ttctatatct gatttgcttg ctattgggcg cggtaatgat
3421 tcctacgatg aaaataaaaa cggcttgctt gttctcgatg agtgcggtac ttggtttaat
3481 acccggttctt ggaatgataa ggaaaagacg ccgattattg attgggttct acatgctcgt
3541 aaattaggat gggatattat ttttcttgtt caggacttat ctattgttga taaacaggcg
3601 cgttctgcat tagctgaaca tggtgtttat tgcgctcgtc tggacagaat tactttacct
3661 tttgtcggta ctttatattc tcttattact ggctcgaaaa tgcctctgcc taaattacat
3721 gttggcggtg ttaaatatgg cgattctcaa ttaagcccta ctgttgagcg ttggctttat
3781 actggtaaga atttgataa cgcatatgat actaaacagg ctttttctag taattatgat
3841 tccgggtgtt attcttattt aacgccttat ttatcacacg gtcggtattt caaacatta
3901 aatttaggtc agaagatgaa attaactaaa atatatattg aaaaagtttc tcgctgtctt
3961 tgccttgca ttggatttgc atcagcattt acatatagtt atataacca acctaagccg
4021 gaggttaaaa aggtagtctc tcagacctat gattttgata aattcactat tgactcttct
4081 cagcgtctta atctaagcta tcgctatggt tcaaggatt ctaagggaat attaatat
4141 agcgacgatt tacagaagca aggttattca ctcacatata ttgatttatg tactgtttcc
4201 attaaaaaag gtaattcaaa tgaaattggt aaatgtaatt aattttgtt tcttgatggt
4261 tgtttcatca tcttcttttg ctcaggtaat tgaaatgaat aattcgctc tcgctgattt
4321 tgtaacttgg tattcaaaagc aatcaggcga atccgttatt gtttctcccg atgtaaaagg
4381 tactgttact gtatattcat ctgacgttaa acctgaaaaa ctacgcaatt tctttattc
4441 tgttttacgt gcaataaatt ttgatatggt aggttctaac ccttcataa ttcagaagta
4501 taatccaaac aatcaggatt atattgatga attgccatca tctgataatc aggaatatga
4561 tgataattcc gctccttctg gtggtttctt tgttccgcaa aatgataatg ttactcaaac
4621 tttttaaatt aataacgttc gggcaaaagga ttaataacga gttgtcgaat tgtttgtaaa
4681 gtctaatact tctaaatcct caaatgtatt atctattgac ggctctaact tattagtgt
4741 tagtgctcct aaagatatat tagataaacct tcctcaattc ctttcaactg ttgatttgcc
4801 aactgaccag atattgattg aggggttgat atttgaggtt cagcaagggt atgctttaga
4861 tttttcattt gctgctggct ctacgcgtgg cactgttgca ggcggtgtta atactgaccg
4921 cctcacctct gttttatctt ctgctggtgg ttcgttcggt atttttaatg gcgatgtttt
4981 agggctatca gttcgcgcac taaagactaa tagccattca aaaatattgt ctgtgccaag
5041 tattcttacg ctttcaggtc agaagggttc tatctctggt ggccagaatg tccctttat
5101 tactgggtcgt gtgactggtg aatctgccaa tgtaataaat ccatttcaga cgattgagcg
5161 tcaaaatgta ggtatttcca tgagcgtttt tcctgttgca atggctggcg gtaatatgt
5221 tctggatatt accagcaagg ccgatatggtt gagttcttct actcaggcaa gtgatgttat
5281 tactaatcaa agaagtattg ctacaacggt taatttgcgt gatggacaga ctcttttact
5341 cgggtggcctc actgattata aaaacacttc tcaggattct ggcgtaccgt tcctgtctaa
5401 aatcccttta atcggcctcc tgtttagctc ccgctctgat tctaacgagg aaagcacgtt
5461 atacgtgctc gtcaaagcaa ccatagtacg cgccctgtag cggcgcatat agcgcggcgg
5521 gtgtggtggt tacgcgcagc gtgaccgcta cacttgccag cgccctagcg cccgctcctt
5581 tcgctttctt ccttctctt ctgcgcacgt tcgcccgtt tccccgtcaa gctctaaatc
5641 gggggctccc tttagggttc cgatttagtg ctttacggca cctcgacccc aaaaaacttg
5701 atttgggtga tggttcacgt agtgggcat cgccctgata gacggttttt cgcccttga
5761 cgttggagtc cacgttcttt aatagtggac tcttgttcca aactggaaca acactcaacc
5821 ctatctcggg ctattctttt gatttataag ggattttgcc gatttcgga ccaccatcaa
5881 acaggatttt cgcctgctgg ggcaaacagg cgtggaccgc ttgctgcaac tctctcaggg
5941 ccaggcgggt aagggcaatc agctgttgcc cgtctcactg gtgaaaagaa aaaccaccct
6001 ggatccaagc ttgcagggtg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta
6061 tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg ataaatgctt
6121 caataatatt gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc ccttattccc
6181 ttttttgagg cattttgcct tcctgttttt gctcaccag aaacgctggt gaaagtaaaa
6241 ttgtctgaag atcagttggg gcactagtgc gggtacatcg aactggatct caacagcggg
6301 aagatccttg agagttttcg cccggaagaa cgttttccaa tgatgagcac ttttaaagtt
6361 ctgctatgtg gcgcgggtat atcccgtatt gacgcggggc aagagcaact cggctgcgcg

```

```

6421 atacactatt ctcagaatga cttggttgag tactcaccag tcacagaaaa gcactcttacg
6481 gatggcatga cagtaagaga attatgcagt gctgccataa ccatgagtga taacactgcg
6541 gccacttac ttctgacaac gatcgaggga ccgaaggagc taaccgcttt tttgcacaac
6601 atgggggata atgtaactcg ccttgatcgt tgggaaccgg agctgaatga agccatacca
6661 aacgacgagc gtgacaccac gatgcctgta gcaatggcaa caacgttcgc caaactatta
6721 actggcgaaac tacttactct agcttcccgg caacaattaa tagactggat ggaggcggat
6781 aaagtgtcag gaccacttct gcgctcggcc cttccggctg gctggtttat tgctgataaa
6841 tctggagcgc gtgagcgtgg gtctcgcggt atcattgcag cactggggcc agatggtaag
6901 ccctcccgtg tcgtagtatt ctacacgacg gggagtcagg caactatgga tgaacgaaat
6961 agacagatcg ctgagatagg tgcctcactg attaagcatt ggtaaactgtc agaccaagtt
7021 tactcatata tactttatgat tgatttaaaa cttcattttt aatttaaaag gatctagggtg
7081 aagatccttt ttgataatct catgaccaa atcccttaac gtgagttttc gtccactgt
7141 acgtaagacc cccaagcttg tcgactgaat ggcgaatggc gctttgcctg gtttccggca
7201 ccagaagcgg tgccggaaag ctggctggag tgcgatcttc ctgacgctcg agcgcaacgc
7261 aattaatgtg agttagctca ctcataggc accccaggct ttacacttta tgcttccggc
7321 tcgtatgttg tgtggaattg tgagcggata acaatttcac acaggaaaca gctatgacca
7381 tgattacgcc aagcttttga gcctttttt tggagatttt caacatgaaa tactattgc
7441 ctacggcagc cgtggattg ttattactcg cGGCCcagcc GGCCatggcc gaagttcaat
7501 tgtagagtc tggtagcgtt cttgttcagc ctggtggttc tttacgtctt tcttgccgtg
7561 cttccggatt cactttctct tcgtacgcta tgtcttgggt tcgccaaact cctggtaaa
7621 gtttggagtg ggtttctgct atctctggtt ctggtggcag tacttactat cctgactccg
7681 ttaaaggctg cttcactatc tctagagaca actctaagaa tactctctac ttgcagatga
7741 acagcttaag ggtgaggac actgcagtct actattgcgc taaagcctat cgtcctctt
7801 atcatgacat atggggtcaa ggtactatgg tcaccgtctc tagtgccctc accaagggcc
7861 catcgggtctt ccgctagca ccctctcca agagcacctc tgggggcaca gcggccctgg
7921 gctgcctggt caaggactac ttcccgaac cgggtgacgt gtctggaac tcaggcgccc
7981 tgacagcgg cgtccacacc ttcccggctg tcctacagtc ctacggactc tactcctca
8041 gcagcgtagt gaccgtgccc tccagcagct tgggcaccca gacctacatc tgcaacgtga
8101 atcacaagcc cagcaacacc aaggtggaca agaaagtga gcccaaatct tgtgcgccc
8161 cacatcatca tcaccatcac ggggcccagc aaaaaaact catctcagaa gaggatctga
8221 atggggccgc agaggctagc tctgctagtg gcgacttcga ctacgagaaa atggttaatg
8281 ccaacaaagg cgcctagact gagaacgctg acgagaatgc tttgcaaaagc gatgccaaag
8341 gtaagttaga cagcgtcgcg accgactatg gcgcgcccat cgacggcttt atcgcgatg
8401 tcagtgggtt ggccaacggc aacggagcca cgggagactt cgcagggttcg aattctcaga
8461 tggcccagggt tggagatggg gacaacagtc cgcttatgaa caactttaga cagtaccttc
8521 cgtctcttcc gcagagtgtc gagtgcgctc cattcgtttt cggtgccggc aagccttacg
8581 agttcagcat cgactgcgat aagatcaatc ttttcggcgg cgttttcgct tcttgctat
8641 acgtcgctac tttcatgtac gttttcagca ctttcgcaa tattttacgc aacaaagaaa
8701 gctagtgtac tcctaggaag ccgcgcta at gagcgggctt tttttttctg gtatgcatcc
8761 tgaggccgat actgtcgtcg tcccctcaaa ctggcagatg cacggttacg atgcgccat
8821 ctacaccaac gtgacctatc ccattacggg caatccgccc tttgttccca cggagaatcc
8881 gacgggttgt tactcgctca catttaatgt tgatgaaagc tggctacagg aaggccagac
8941 gcgaattatt tttgatggcg ttcctattgg ttaaaaaatg agctgattta acaaaaaatt
9001 aatgcgaatt ttaacaaaat attaacgttt acaatttaaa tatttgctta tacaatcttc
9061 ctgttttttg ggcttttctg attatcaacc ggggtacata tgattgacat gctagtttta
9121 cgattaccgt tcattcgatc tcttgtttgc tccagactct caggcaatga cctgatagcc
9181 tttgtagatc tctcaaaaat agctaccctc tccggcatta atttatcagc tagaacgggt
9241 gaatatcata ttgatgggtg tttgactgtc tccggccttt ctacccttt tgaatcttta
9301 cctacacatt actcaggcat tgcatttaaa atatatgagg gttctaaaaa tttttatcct
9361 tgcgttgaaa taaaggcttc tcccgcaaaa gtattacagg gtcataatgt ttttgggtaca
9421 accgatttag ctttatgctc tgaggcttta ttgcttaatt ttgctaattc tttgccttgc
9481 ctgtatgatt tattggatgt t

```

Tabla 35: Secuencia de ADN de pMID21: 5957 pb (SEQ ID NO:895)

1 gacgaaaggg cctcgtgata cgcctatttt tatagggttaa tgcctatgata ataattggtt

```

61  cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aaccctatt tgtttatatt
121 tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat
181 aatattgaaa aaggaagagt atgagtattc aacatttccg tgcgcacctt attccctttt
241 ttgcggcatt ttgccttcct gtttttgctc acccagaaaac gctggtgaaa gtaaaagatg
301 ctgaagatca gttgggtgcc cgagtgggtt acatcgaact ggatctcaac agcggtaaga
361 tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagtctctg
421 tatgtggcgc ggtattatcc cgtattgacg cggggcaaga gcaactcggc cgccgcatac
481 actatttctc gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg
541 gcatgacagt aagagaatta tgcagtgtcg ccataaccat gagtgataac actgcgccca
601 acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg
661 gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg
721 acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaaactg
781 gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggaataaa
841 ttgcaggacc acttctgcgc tccggccttc cggctggctg gtttattgct gataaaactg
901 gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct
961 cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac
1021 agatcgctga gatagggtcc tcactgatta agcattggta actgtcagac caagtttact
1081 catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga
1141 tcctttttga taatctcatg accaaaatcc cttaacgtga gtttctgttc cactgagcgt
1201 cagaccccggt agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct
1261 gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccc gatcaagagc
1321 taccactctt ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgttc
1381 ttctagtgtg gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc
1441 tcgctctgct aatcctgta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg
1501 ggttggaactc aagacgtatg ttaccggata aggcgcagcg gtcgggctga acggggggtt
1561 cgtgcataca gccagccttg gagcgaacga cctacaccga actgagatac ctacagcgtg
1621 agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggttaagcg
1681 gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgct ttggtatctt
1741 atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgcctctcag
1801 gggggcgagg cctatggaaa aacgcagca acgcgccctt ttacggttc ctggcctttt
1861 gctggccttt tgctcacatg ttcttctctg cgttatcccc tgattctgtg gataaccgta
1921 ttaccgcctt tgagtgaact gataccgttc gccgcagccg aacgaccgag cgcagcaggt
1981 cagtgcagca ggaagcggaa gagcgcccaa tacgcaaac gcctctcccc gcgcgttggc
2041 cgattcatta atgcagctgg cagcagaggt tcccgaactg gaaagcgggc agtgagcgca
2101 acgcaattaa gtgagtttag ctcaactcatt aggcacccca ggctttacac tttatgcttc
2161 cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg
2221 accatgatta cgccaagcct tggagccttt tttttggaga ttttcaacgt gaaaaaatta
2281 ttattcgcaa ttcttttagt tgttcctttc tattctcaca gtgcacaggt ccaactgcag
2341 gagctcgaga tcaaacgtgg aactgtggct gcaccatctg tcttcatctt cccgccatct
2401 gatgagcagt tgaatctgg aactgcctct gttgtgtgcc tgctgaataa ctctatccc
2461 agagaggcca aagtacagtg gaaggtggat aacgcctcc aatcgggtaa ctcccaggag
2521 agtgcacag agcaggacag caaggacagc acctacagcc tcagcagcac cctgacgctg
2581 agcaaaagcag actacgagaa acacaaagtc tacgcctgcg aagtcaccca tcaggccctg
2641 agttcaccgg tgacaaagag cttcaacagg ggagagtgtt aataagcgc gcctaaccat
2701 ctatttcaag gaacagctct aatgaaaaag cttttattca tgatcccgtt agttgtaccg
2761 ttctgtggccc agccggcctc tgctgaagtt caattgttag agtctggtgg cggctctgtt
2821 cagcctggtg gttctttacg tctttcttgc gctgcttccg gagcttcaga tctgtttgcc
2881 tttttgtggg gtggtgcaga tcgcgttacg gagatcgacc gactgcttga gcaaaaagcca
2941 cgcttaactg ctgatcaggc atgggatgtt attcgccaaa ccagtcgtca ggatcttaac
3001 ctgagggcttt ttttacctac tctgcaagca gcgacatctg gtttgacaca gagcgatccg
3061 cgtcgtcagt tggtagaaac attaacacgt tgggatggca tcaatttgct taatgatgat
3121 ggtaaaaacct ggcagcagcc aggctctgcc atcctgaacg tttggctgac cagtatgttg
3181 aagcgtaccg tagtggctgc cgtacctatg ccatttgata agtggtagag cgccagtggc
3241 tacgaacaa cccaggacgg cccaactggt tcgctgaata taagtgttgg agcaaaaatt
3301 ttgtatgagg cgggtgcagg agacaaatca ccaatcccac aggcggttga tctgtttgct
3361 gggaaaccac agcaggaggt tgtgttggt gcgctggaag atacctggga gactctttcc
3421 aaacgctatg gcaataatgt gagtaactgg aaaaaccagg caatggcctt aacgttccgg

```

```

3481 gcaataaatt tctttggtgt accgcaggcc gcagcggaag aaacgcgtca tcaggcggag
3541 tatcaaaacc gtggaacaga aaacgatatg attgttttct caccaacgac aagcgatcgt
3601 cctgtgcttg cctgggatgt ggtcgacacc ggtcagagtg ggtttattgc tcccgatgga
3661 acagtttgata agcactatga agatcagctg aaaatgtacg aaaattttgg ccgtaagtcg
3721 ctctggttaa cgaagcagga tgtggaggcg cataaggagt tctagagaca actctaagaa
3781 tactctctac ttgcagatga acagcttaag tctgagcatt cgtccggggc aacattctcc
3841 aaactgacca gacgacacaa acggcttacg ctaaatcccg cgcattggat ggtaaaaggg
3901 tggcgctctt gctggccttg actcatcaga tgaaggccaa aaattggcag gagtggacac
3961 agcaggcagc gaaacaagca ctgaccatca actggtacta tgctgatgta aacggcaata
4021 ttggttatgt tcatactggt gcttatccag atcgtcaatc aggccatgat ccgcgattac
4081 ccgttccttg tacgggaaaa tgggactgga aagggtctatt gccttttgaa atgaacccta
4141 aggtgtataa cccccagcag ctagccatat tctctcggtc accgtctcaa gcgcctccac
4201 caagggccca tcggtcttcc cgctagcacc ctctccaag agcacctctg ggggcacagc
4261 ggccctgggc tgctgtgta aggaactact cccggaaccg gtgacggtg cgtggaactc
4321 aggcgccttg accagcggcg tccacacett cccggtgtc ctacagtcta gcggactcta
4381 ctccctcagc agcgtagtga ccgtgccttc ttctagcttg ggcacccaga cctacatctg
4441 caacgtgaat cacaagccca gcaacaccaa ggtggacaag aaagttgagc ccaaattctg
4501 tgccgcccga catcatcac accatcacgg ggccgcagaa caaaaactca tctcagaaga
4561 ggatctgaat ggggccgcag aggctagttc tgctagtaac gcgtcttcg gtgattttga
4621 ttatgaaaag atggcaaacg ctaataaggg ggctatgacc gaaaatgccg atgaaaacgc
4681 gctacagtct gacgctaaag gcaaaactga ttctgtcgct actgattacg gtgctgctat
4741 cgatggtttc attggtgacg tttccggcct tgctaattgt aatggtgcta ctggtgattt
4801 tgctggctct aattcccaaa ttggtcaagt cggtagcggg gataattcac ctttaatgaa
4861 taatttccgt caatatttac ctccctccc tcaatcggtt gaatgtcgcc tttttgtctt
4921 tggcgctggt aaaccatatg aattttctat tgattgtgac aaaataaact tattccgtgg
4981 tgtctttgcg tttcttttat atgttgccac ctttatgtat gtattttcta cgtttgctaa
5041 catactgctg aataaggagt cttaatgaaa cgcgtgatga gaattcactg gccgtcgttt
5101 tacaacgctg tgactgggaa aaccctggcg ttaccaact taatcgctt gcagcacatc
5161 cccctttcgc cagctggcgt aatagcgaag aggcgcgcac cgatcgccct tcccaacagt
5221 tgccgagcct gaattggcgaa tggcgctga tgcggtatct tctccttacg catctgtgcy
5281 gtatttcaca ccgcatacgt caaagcaacc atagtacgcg ccctgtagcg gcgcattaag
5341 cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccttagcgcc
5401 cgctcctttc gctttcttcc ctcccttct cgccacgttc gccggcttc ccgctcaagc
5461 tctaaatcgg gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa
5521 aaaacttgat ttgggtgatg gttcacgtag tgggccatcg ccctgataga cgttttttcg
5581 ccctttgacg ttggagtcca cgttctttaa tagtggaactc ttgttccaaa ctggaacaac
5641 actcaactct atctcgggct attcttttga ttataaaggg attttgccga tttcgggtcta
5701 ttggttaaaa aatgagctga tttaacaaaa atttaacgcg aattttaaca aaatattaac
5761 gtttacaatt ttatggtgca gtctcagtag aatctgctct gatgccgat agttaagcca
5821 gcccgcacac ccgccaacac ccgtgacgc gccctgacgg gcttgtctgc tcccggcatc
5881 cgcttacaga caagctgtga ccgtctccgg gagctgcatg tgtcagaggt tttcaccgctc
5941 atcaccgaaa cgcgcga

```

Tabla 36: pM21J que contiene Ills::A27::Ckappa

Número de bases 5225 (SEQ ID N:921)

```

GACGAAAGGG CCICGTGATA CGCCTATTTT TATAGGTTAA TGTCATGATA ATAATGGTTT 60
CTTAGACGTC AGGTGGCACT TTTCGGGGAA ATGTGCGCGG AACCCCTATT TGTTTATTTT 120
TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA ATGCTTCAAT 180
AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT 240
TTGCGGCATT TTGCCTTCCT GTTTTIGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG 300
CTGAAGATCA GTTGGGTGCC CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA 360
TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC 420
TATGTGGCGC GGTATTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC 480
ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG 540
GCATGACAGT AAGAGAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA 600

```

ACTTACTTCT	GACAACGATC	GGAGGACCGA	AGGAGCTAAC	CGCTTTTTTG	CACAACATGG	660
GGGATCATGT	AACTCGCCTT	GATCGTTGGG	AACCGGAGCT	GAATGAAGCC	ATACCAAACG	720
ACGAGCGTGA	CACCACGATG	CCTGTAGCAA	TGGCAACAAC	GTTGCGCAAA	CTATTAACTG	780
GCGAACTACT	TACTCTAGCT	TCCCGGCAAC	AATTAATAGA	CTGGATGGAG	GCGGATAAAG	840
TTGCAAGGACC	ACTTCTGCGC	TCGGCCCTTC	CGGCTGGCTG	GTTTATTGCT	GATAAATCTG	900
GAGCCGGTGA	GGGTGGGTCT	CGCGGTATCA	TTGCAGCACT	GGGGCCAGAT	GGTAAGCCCT	960
CCCGTATCGT	AGTTATCTAC	ACGACGGGGA	GTCAAGCAAC	TATGGATGAA	CGAAATAGAC	1020
AGATCGCTGA	GATAGGTGCC	TCACTGATTA	AGCATTGGTA	ACTGTCAGAC	CAAGTTTACT	1080
CATATATACT	TTAGATTGAT	TTAAAACTTC	ATTTTAAATT	TAAAAGGATC	TAGGTGAAGA	1140
TCCTTTTTGA	TAATCTCATG	ACCAAAATCC	CTTAACGTGA	GTTTTCGTTC	CACTGAGCGT	1200
CAGACCCCGT	AGAAAAGATC	AAAGGATCTT	CTTGAGATCC	TTTTTTTCTG	CGCGTAATCT	1260
GCTGCTTGCA	AACAAAAAAA	CCACCGCTAC	CAGCGGTGGT	TTGTTTGCCG	GATCAAGAGC	1320
TACCAACTCT	TTTTCCGAAG	GTAAGTGGCT	TCAGCAGAGC	GCAGATACCA	AATACTGTTC	1380
TTCTAGTGTA	GCCGTAGTTA	GGCCACCAC	TCAGAACTC	TGTAGCACCG	CCTACATACC	1440
TCGCTCTGCT	AATCCTGTTA	CCAGTGGCTG	CTGCCAGTGG	CGATAAGTCG	TGCTTTACCG	1500
GGTTGGACTC	AAGACGATAT	TTACCGGATA	AGGCGCAGCG	GTGCGGCTGA	ACGGGGGGGT	1560
CGTGCATACA	GCCCAGTTTG	GAGCGAACGA	CCTACACCGA	ACTGAGATAC	CTACAGCGTG	1620
AGCTATGAGA	AAGCGCCACG	CTTCCCGAAG	GGAGAAAGGC	GGACAGGTAT	CCGGTAAGCG	1680
GCAGGGTCGG	AACAGGAGAG	CGCAGCAGGG	AGCTTCCAGG	GGGAAACGCC	TGGTATCTTT	1740
ATAGTCTCTG	CGGGTTTCGC	CACCTCTGAC	TTGAGCGTCG	ATTTTGTGTA	TGCTCGTCAG	1800
GGGGGCGGAG	CCTATGGAAA	AACGCCAGCA	ACGCGGCCCT	TTTACGGTTC	CTGGCCTTTT	1860
GCTGGCCCTT	TGCTCACATG	TTCTTTCCTG	CGTTATCCCC	TGATTCTGTG	GATAACCGTA	1920
TTACCGCCCT	TGAGTGAGCT	GATACCGCTC	GCCGCAGCCG	AACGACCGAG	CGCAGCGAGT	1980
CAGTAGCGGA	GGAAGCGGAA	GATCGCCCAA	TACGCAAAAC	GCCTCTCCCC	GCGCGTTGGC	2040
CGATTCAATTA	ATGCAGCTGG	CACGACAGGT	TTCCCGACTG	GAAAGCGGGC	AGTGAGCGCA	2100
ACGCAATTA	TGTGAGTTAG	CTCACTCAT	AGGCACCCCA	GGCTTTTACAC	TTTATGCTTC	2160
CGGCTCGTAT	GTTGTGTGGA	ATTGTGAGCG	GATAACAATT	TCACACAGGA	AACAGCTATG	2220
ACCATGATTA	GCTCAGCTTT	TGGAGCCCTT	TTTTTGGAGA	TTTTCAACAT	GAAGAAACTG	2280
CTGTCTGCTA	TCCCACATAG	TGTCCTTTTC	TATCTCTATA	GTGAAATCGT	TCTGACCCAG	2340
TCCCGGGGGA	CCCTGTCTCT	GTCTCCGGGT	GAACGTGCTA	CGCTGAGCTG	TCGTGCTTCT	2400
CAATCCGTTA	GCTCCTCTTA	TTTAGCTTGG	TATCAGCAAA	AGCCGGGTCA	AGCTCCGCGG	2460
CTGTTGATCT	ATGGTGCTCT	TAGTCGTGCT	ACTGGCATCC	CTGATCGTTT	CTCTGGCTCT	2520
GGCTCCGGAA	CCGATTTTAC	TCTGACCAAT	TCTCGTCTCG	AGCCGGAAGA	TTTCGCTGTC	2580
TACTATTGTC	AACAGTATGG	TCTAGTCCG	CTGACTTTTC	GTGGCGGTAC	CAAAGTCGAA	2640
ATCAAGCGTG	GAAGTGTGGC	TGCACCATCT	GTCTTCATCT	TCCCGCCATC	TGATGAGCAG	2700
TTGAAATCTG	GAAGTGTGGC	TGTTGTGTGC	CTGCTGAATA	ACTTCTATCC	CAGAGAGGCC	2760
AAAGTACAGT	GGAAGGTGGA	TAACGCCCTC	CAATCGGGTA	ACTCCAGGGA	GAGTGTCACA	2820
GAGCAGGACA	GCAAGGACAG	CACCTACAGC	CTCAGCAGCA	CCCTGACTCT	GTCCAAAGCA	2880
GACTACGAGA	AACACAAAGT	CTACGCCTGC	GAAGTCACCC	ATCAGGGCCT	GAGTTCACCG	2940
GTGACAAAGA	GCTTCAACAG	GGGAGAGTGT	TAATAAGGCG	CGCCAATTTA	ACCATCTATT	3000
TCAAGGAACA	GCTTAAATGA	AGAAGCTCCT	CTTTGCTATC	CCGCTCGTCG	TTCCTTTTGT	3060
GGCCAGCCGG	GCCATGGCCG	AAGTTCAATT	GTTAGAGTCT	GGTGGCGGTC	TGTTTCAGCC	3120
TGGTGGTTCT	TTACGTCCTT	CTTGCCTGTC	TTCCGGATTG	ACTTTCTCTC	GTTACAAGAT	3180
GAAGTGGGTT	GCCCAAGCTC	CTGGTAAAGG	TTTGGAGTGG	GTTTCTGTTA	TCTATCCTTC	3240
TGGTGGCGGT	ACTGGTTATG	CTGACTCCGT	TAAAGGTCGC	TTCACTATCT	CTAGAGACAA	3300
CTCTAAGAAT	ACTCTCTACT	TGCAGATGAA	CAGCTTAAGG	GCTGAGGACA	CTGCAGTCTA	3360
CTATTGTGCG	AGAGTCAATT	ACTATGATAG	TAGTGGTTAG	GGTCTATAG	CTCCTGGACT	3420
TGACTACTGG	GGCCAGGGAA	CCCTGGTCAC	CGTCTCAAGC	GCCTCCACCA	AGGGTCCGTC	3480
GGTCTTCCCG	CTAGCACCTT	CCTCCAAGAG	CACCTCTGGG	GGCACAGCGG	CCCTGGGCTG	3540
CCTGGTCAAG	GACTACTTCC	CCGAACCGGT	GACGGTGTGC	TGGAAGTCAG	GCGCCCTGAC	3600
CAGCGGCGTC	CACACTTCC	CGGCTGTCTC	ACAGCTTAGC	GGACTCTACT	CCCTCAGCAG	3660
CGTAGTGACC	GTGCCCTCTT	CTAGCTTGGG	CACCCAGACC	TACATCTGCA	ACGTGAATCA	3720
CAAGCCCAGC	AACACCAAGG	TGGACAAGAA	AGTTGAGCCC	AAATCTTTGT	CGGCCGCACA	3780
TCATCATCAC	CATCAGGGGG	CCGAGAACA	AAAACCTCAT	TCAGAAGAGG	ATCTGAATGG	3840
GGCCGACAG	GCTAGTTCTG	CTAGTAACGC	GTCTTCCGGT	GATTTTGATT	ATGAAAAGAT	3900
GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AAATGCCGAT	GAAAACGCGC	TACAGTCTGA	3960
CGCTAAAGGC	AACTTGATT	CTGTCGCTAC	TGATTACGGT	GCTGCTATCG	ATGGTTTCAT	4020
TGGTGACGTT	TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	GGTGATTTTG	CTGGCTCTAA	4080
TTCCCAAATG	GCTCAAGTCG	GTACGGGTGA	TAATTCACCT	TAAATGAATA	ATTTCGGTCA	4140
ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT	TTTGTCTTTG	GCGCTGGTAA	4200
ACCATATGAA	TTTCTATTG	ATTGTGACAA	AATAAACTTA	TTCCGTGGTG	TCTTTGCGTT	4260
TCTTTTATAT	GTTGCCACCT	TTATGTATGT	ATTTTCTAGC	TTTGCTAACA	TACTGCGTAA	4320
TAAGGAGTCT	TAATGAAACG	CGTGATGAGA	ATTCACTGGC	CGTCTGTTTA	CAACGTCGTG	4380
ACTGGGAAAA	CCCTGGCGTT	ACCCAACCTA	ATCGCCTTGC	AGCACATCCC	CCTTTGCCCA	4440
GCTGGCGTAA	TAGCGAAGAG	CGCCGCACCG	ATCGCCCTTC	CCAACAGTTG	CGCAGCCTGA	4500
ATGGCGAATG	GCGCCTGATG	CGGTATTTTC	TCCTTACGCA	TCTGTGCGGT	ATTTACACAC	4560
GCATACGTCA	AAGCAACCAT	AGTACGCGCC	CTGTAGCGGC	GCATTAAAGC	CGGCGGGTGT	4620
GGTGGTTACG	CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	TAGCGCCCG	CTCCTTTCCG	4680
TTTCTTCCCT	TCCTTTCTCG	CCACGTTCCG	CGGCTTTCCC	CGTCAAGCTC	TAAATCGGGG	4740
GCTCCCTTTA	GGGTTCCGAT	TTAGTGCTTT	ACGGCACCTC	GACCCCAAAA	AACTTGATT	4800
GGGTGATGGT	TCACGTAGTG	GGCCATCGCC	CTGATAGACG	GTTTTTCGCC	CTTTGACGTT	4860
GGAGTCCACG	TTCTTTAATA	TGGAGCTCTT	GTTCCAAACT	GGAACAACAC	TCAACTCTAT	4920
CTCGGGCTAT	TCTTTTGATT	TATAAGGGAT	TTTGCCGATT	TGGGTCTATT	GGTTAAAAAA	4980
TGAGCTGATT	TAACAAAAAT	TTAACGCGAA	TTTTAACAAA	ATATTAAAGT	TTACAATTTT	5040
ATGGTGACGT	CTCAGTACAA	TCTGCTCTGA	TGCCGCATAG	TTAAGCCAGC	CCCGACACCC	5100
GCCAAACCCC	GCTGACGCGC	CCTGACGGGC	TTGCTGTGTC	CCGGCATCCG	CTTACAGACA	5160
AGCTGTGACC	GTCTCCGGGA	GCTGCATGTG	TCAGAGGTTT	TCACCGTCAT	CACCGAAACG	5220
CGCGA						5225

Tabla 40: pLCSK23 (SEQ ID NO:896)

1	GACGAAAGGG	CCTGCTCTGC	CAGTGTAC	ACCAATTAAC	CAATTCTGAT	TAGAAAACT
61	CATCGAGCAT	CAATGAAAC	TGCAATTTAT	TCATATCAGG	ATTATCAATA	CCATATTTTT
121	GAAAAAGCCG	TTTCTGTAAT	GAAGGAGAAA	ACTCACCAG	GCAGTTCCAT	AGGATGGCAA
181	GATCCTGGTA	TCGGTCTGCG	ATTCCGACTC	GTCCAACATC	AATACAACCT	ATTAATTTCC
241	CCTCGTCAAA	AATAAGGTTA	TCAAGTGAGA	AATCACCATG	AGTGACGACT	GAATCCGGTG
301	AGAATGGCAA	AAGCTTATGC	ATTTCTTTCC	AGACTTGTTC	AACAGGCCAG	CCATTACGCT
361	CGTCATCAAA	ATCACTCGCA	TCAACCAAAC	CGTTATTTCAT	TCGTGATTGC	GCCTGAGCGA
421	GACGAAATAC	GCGATCGCTG	TAAAAGGAC	AATTACAAAC	AGGAATTGAA	TGCAACCGGC
481	GCAGGAACAC	TGCCAGCGCA	TCAACAATAI	TTTCACCTGA	ATCAGGATAT	TCTTCTAATA
541	CCTGGAATGC	TGTTTTCCCG	GGGATCGCAG	TGGTGAGTAA	CCATGCATCA	TCAGGAGTAC
601	GGATAAAATG	CTTGATGGTC	GGAGAGGCA	TAAATCCCGT	CAGCCAGTTT	AGTCTGACCA
661	TCTCATCTGT	AACATCATTTG	GCAACGCTAC	CTTTGCCATG	TTTCAGAAAC	AACTCTGGCG
721	CATCGGGCTT	CCCATACAAT	CGATAGATTG	TCGCACCTGA	TTGCCCGACA	TTATCGCGAG
781	CCCATTITATA	CCCATATAAA	TCAGCATCCA	TGTTGGAATT	TAATCGCGCG	CTCGAGCAAG
841	ACGTTTCCCG	TGGAATATGG	CTCATAACAC	CCCTTGATAT	ACTGTTTATG	TAAGCAGACA
901	GTTTTATTGT	TCATGATGAT	ATATTTTTAT	CTTGTGCAAT	GTAACATCAG	AGATTTTGAG
961	ACACAACGTG	GCTTTCCCCC	CCCCCCCCTG	CAGGTCTCGG	GCTATTCTTG	TCAGACCAAG
1021	TTTACTCATA	TATACTTTAG	ATTGATTTAA	AACTTCATTT	TTAATTTAAA	AGGATCTAGG
1081	TGAAGATCGT	TTTTGATAAT	CTCATGACCA	AAATCCCTTA	ACGTGAGTTT	TCGTTCCACT
1141	GAGCGTCAGA	CCCGGTAGAA	AAGATCAAAG	GATCTTCTTG	AGATCCTTTT	TTTCTGCGCG
1201	TAATCTGCTG	CTTGCAAAAC	AAAAAACCCAC	CGCTACCAGC	GGTGGTTTGT	TTGCCGGATC
1261	AAGAGCTACC	AACTCTTTTT	CCGAAGGTAA	CTGGCTTCAG	CAGAGCGCAG	ATACCAAATA
1321	CTGTTCTTCT	AGTGTAGCCG	TAGTTAGGCC	ACCACTTCAA	GAACCTCTGA	GCACCCGCTA
1381	CATACCTCGC	TCTGCTAATC	CTGTTACCAG	TGGCTGCTGC	CAGTGGCGAT	AAGTCGTGTC
1441	TTACCGGGT	GGACTCAAGA	CGATAGTTAC	CGGATAAGGC	GCAGCGTTCG	GGCTGAACGG
1501	GGGGTTCGTG	CATACAGCCC	AGCTTGGAGC	GAACGACCTA	CACCGAAGTG	AGATACCTAC
1561	AGCGTGAGCT	ATGAGAAAGC	GCCACGCTTC	CCGAAGGGAG	AAAGGCGGAC	AGGTATCCGG
1621	TAAGCGGCAG	GGTCGGAACA	GGAGAGCGCA	CGAGGGAGCT	TCCAGGGGGA	AACGCCCTGGT
1681	ATCTTTATAG	TCTGTGCGGG	TTTCGCCACC	TCTGACTTGA	GCGTCGATTT	TTGTGATGCT
1741	CGTCAGGGGG	GCGGAGCCTA	TGGAAAAACG	CCAGCAACGC	GGCCTTTTTA	CGGTTCTCTG
1801	CCTTTTGCTG	GCCTTTTGCT	CACATGTTCT	TTCTTGCGTT	ATCCCCTGAT	TCTGTGGATA
1861	ACCGTATTAC	CGCCTTTGAG	TGAGCTGATA	CCGCTCGCCG	CAGCGGAACG	ACCGAGCGCA
1921	GCGAGTCAGT	GAGCGAGGAA	GCGGAAGAGC	GCCCAATACG	CAAACCGCCT	CTCCCCGCGC
1981	GTGGGCCGAT	TCATTAATGC	AGCTGGCAGC	ACAGGTTTCC	CGACTGGAAA	GCGGGCAGTG
2041	AGCGCAACGC	AATTAATGTG	AGTTAGCTCA	CTCATTAGGC	ACCCAGGCT	TTACACTTTA
2101	TGCTTCCGGC	TCGTATGTTG	TGTGGAATTG	TGAGCGGATA	ACAATTTTCA	ACAGGAAACA
2161	GCTATGACCA	TGATTACGCC	AAGCTTTGGA	GCCTTTTTTT	TGGAGATTTT	CAACATGAAG
2221	AAGCTCCTCT	TTGCTATCCC	GCTCGTCGTT	CCTTTTGTGG	CCCAGCCGGC	CATGGCCGAC
2281	ATCCAGATGA	CCCACTCTCC	ATCCTCCCTG	TCTGCATCTG	TAGGAGACAG	AGTCACCATC
2341	ACTTGCCGGG	CAAGTCAGAG	CATTAGCAGC	TATTTAAATT	GGTATCAGCA	GAAACCAAGG
2401	AAAGCCCTTA	AGCTCTCGAT	CTATGCTGCA	TCCAGTTTGC	AAAGTGGGGT	CCCATCAAGG
2461	TTCACTGGCA	GTGGATCTGG	GACAGATTTT	ACTCTCACCA	TCAGCAGTCT	GCAACCTGAA
2521	GATTTTGCAA	CTTACTACTG	TCAACAGAGT	TACAGTACCC	CTTTCACTTT	CGGCCCTGGG
2581	ACCAAAGTGG	ATATCAAACG	TGGTACCGTG	GCTGCACCAT	CTGTCTTCAT	CTTCCCGCCA
2641	TCTGATGAGC	AGTTGAAATC	TGGAACCTGCC	TCTGTTGTGT	GCCTGCTGAA	TAACCTCTAT
2701	CCCAGAGAGG	CCAAAGTACA	GTGGAAGGTG	GATAACGCCC	TCCAATCGGG	TAACCTCCAG
2761	GAGAGTGTC	CAGAGCAGGA	CAGCAAGGAC	AGCACCTACA	GCCTCAGCAG	CACCCGTAGC
2821	CTGAGCAAAG	CAGACTACGA	GAAACACAAA	GTCTACGCTT	GCGAAGTCAC	CCATCAGGGC
2881	CTGAGTTTAC	CGGTGACAAA	GAGCTTCAAC	AGGGGAGAGT	GTGCGGCCGC	TGGTAAGCCT
2941	ATCCCTAAC	CTCTCCCTCG	TCTCGATTCT	ACGTGATAAC	TTACCCGGTC	AACCGCTGAT
3001	GAGAATTAC	TGGCCGTCGT	TTTACAACGT	CGTGACTGGG	AAAACCTTGG	CGTTACCCAA
3061	CTTAATCGCC	TTGCAGCACA	TCCCCCTTTC	GCCAGCTGGC	GTAATAGCGA	AGAGGCCCGC
3121	ACCGATCGCC	CTTCCCAACA	GTTGCGCAGC	CTGAATGGCG	AATGGCGCCT	GATGCGGTAT
3181	TTTCTCCTTA	CGCATCTGTG	CGGTATTTCA	CACCGCATAC	GTCAAAGCAA	CCATAGTCTC
3241	AGTACAATCT	GCTCTGATGC	CGCATAGTTA	AGCCAGCCGC	GACACCCGCG	AACACCCGCT
3301	GACGCGCCCT	GACAGGCTTG	TCTGCTCCCG	GCATCCGCTT	ACAGACAAGC	TGTGACCGTC
3361	TCCGGGAGCT	GCATGTGTCA	GAGGTTTTCA	CCGTCATCAC	CGAAACGCGC	GA

Ejemplo 4: Titubeo digital de CDRs

- 5 Los siguientes ejemplos ejemplifican el uso de titubeo a la hora de construir bibliotecas sintéticas. La cadena pesada (HC) 3-23 progenitora se diversifica en CDR1, 2, y 3. Esta diversidad se combina con una cadena ligera (LC) A27 sintéticamente diversificada. La diversidad será como sigue:

Ejemplo 4.1 CDR1 de HC

- 10 La siguiente diversidad mediante titubeo permite 5.832 variantes. Véase la Tabla 50. En la posición 31, Ser es el tipo de aminoácido de línea germinal (GL). Por tanto, obtuvimos Ser tres veces más probable que los otros tipos. Puesto que se permiten 18 tipos, Ser se permitirá 15% de las veces, y los otros se permitirán 5%. De este modo, si no hay selección para el tipo de AAs en 31, muy probablemente aislaremos un Ab con Ser. De forma similar, en 33, el tipo de AA de GL es Ala, y obtenemos Ala 3 veces igual de probable (15%) que todos los otros (5%). En 35, Ser es el tipo de AA de GL, y lo obtenemos tres veces igual de probable que los otros. En las tres posiciones, se han excluido
- 15 Cys y Met. Se excluye Cys debido a que no se desean disulfuros innecesarios o cisteínas no emparejadas expuestas que podrían afectar de forma adversa a la solubilidad y reactividad del Ab. Se excluye Met debido a que

los grupos laterales de metioninas expuestos están sometidos a oxidación, lo que puede alterar a las propiedades de unión y a la semivida. Se podría obtener el tipo de aminoácido de línea germinal 2, 3, 4, 5, 6, 8, ó 10 veces más probable que los otros tipos de AAs.

Tabla 50: Diversidad para CDR1 en 3-23

Posición	AA progenitor	Permitido
31	S (tres veces más probable que los otros)	ADEFGHKLNPQRSTVWY
33	A (3-X más probable)	ADEFGHKLNPQRSTVWY
35	S (3-X más probable)	ADEFGHKLNPQRSTVWY

A lo largo de esta descripción, los aminoácidos "Permitidos" mostrados son los aminoácidos que se pueden usar en una posición dada. Por ejemplo, en la Tabla 50, en la posición 31, se muestran los aminoácidos permitidos "ADEFGHKLNPQRSTVWY". Esto indica que los aminoácidos A, D, E, F, G, H, K, L, N, P, Q, R, S, T, V, W, e Y están todos permitidos en la posición 31.

Ejemplo 4.2: CDR2 de HC

En CDR2, se permite diversidad en las posiciones 50, 52, 52a, 56, y 58. En 50, 52, 56, y 58 se permiten todos los tipos de aminoácidos excepto Cys y Met, y obtenemos los tipos de AAs de GL más probablemente unas tres veces. Podríamos obtener el tipo de AA de GL 2, 3, 4, 5, 6, 8, ó 10 veces más probable que los otros tipos de AAs.

Tabla 51: CDR2 de HC: Diversidad = 419.904

Posición	AA progenitor	Permitido
50	A (3-X más probable)	ADEFGHKLNPQRSTVWY
52	S (3-X más probable)	ADEFGHKLNPQRSTVWY
52a	G (3-X más probable)	GPSY
56	S (3-X más probable)	ADEFGHKLNPQRSTVWY
58	Y (3-X más probable)	ADEFGHKLNPQRSTVWY

Diversidad combinada de CDR1 y CDR2 = 2,45 E 9.

Ejemplo 4.3: CDR3 de HC, longitudes 3, 4, 5

Las CDR3 muy cortas se pueden obtener mediante titubeo. La Tabla 7 muestra varias secuencias progenitoras para CDR3 de longitud 3. En 94, muchos VH3s tienen Arg y se ha permitido este cambio, pero se obtuvo Lys 3-X igual de probable. En 95, F se encuentra en esta posición en JH1. También se permiten Ser, Tyr, Asp, y Arg para permitir carga positiva pequeña, grande, y carga negativa. En 96, JH1 tiene Q. Puesto que Q es muy similar a Glu, se permite Glu como una alternativa ácido más Arg, Ser, Tyr, y Leu. En 97, His es el AA de línea germinal de JH1. Se permite la carga negativa (D), la carga positiva (R), polar pequeña (S), hidrófobo grande (Y), y alifático (L). La secuencia progenitora constituye el 4,5% de la biblioteca, pero esto se combina con una gran diversidad en CDR1 y CDR2. El titubeo permite 360 secuencias en total. Las secuencias menos probables se producen 1 en 1792. La secuencia más probable (progenitora) se produce alrededor de 1 en 22.

Tabla 60: Una CDR3 de HC titubeada de longitud 3 (V-3JH1 de la Tabla 7)

Posición	Aminoácido progenitor (fuente)	Permitido
94	K (VH 3-23)	KR (3:1)
95	F (JH1)	FSYDR (3:1:1:1:1)
96	Q (JH1)	QERSYL (3:1:1:1:1:1)
97	H (JH1)	HDRSYL (3:1:1:1:1:1)

Posición	Aminoácido progenitor (fuente)	Permitido
103	W (JH1)	W

La Tabla 61 muestra una CDR3 de HC titubeada de longitud 3. Aquí, K94 se fija como lo hace W103. Hemos obtenido el aminoácido del segmento D "progenitor" cinco veces igual de probable que los otros tipos de AAs permitidos.

5 Tabla 61: Una CDR3 de HC titubeada de longitud 3 de un fragmento D (V-3D1-1.1.2-JH1 de la Tabla 7).

Posición	Parental	Permitido
94	K (V 3-23)	K
95	T (D1-1.1.2)	TYRDL (5:1:1:1:1)
96	T (D1-1.1.2)	TYRDL (5:1:1:1:1)
97	G (D1-1.1.2)	GSYRDL (5:1:1:1:1:1)
103	W (JH1)	W

En este ejemplo (Tabla 62, usando V-4JH2 de la Tabla 8), 94 se fija como Lys. En 95, JH2 tiene Tyr, y se han permitido Ser, Asp, Arg, y Leu de manera que el tamaño, la carga y la hidrofobia se pueden alterar para adecuarse al antígeno. JH2 tiene Phe en 96, y hemos permitido Ser, Tyr, Asp, Arg, y Leu. En 97, JH2 tiene Asp, y hemos permitido Arg, Ser, Tyr, y Leu. En 98, JH2 tiene Leu, y hemos permitido Ser, Tyr, Asp, y Arg. Este patrón permite 750 secuencias distintas, de las cuales la progenitora es la más probable (1 en 18). Las secuencias menos probables se producen en 1 en 4608, o son 256 veces menos probables que la más probable.

10 Tabla 62: CDR3 de HC de longitud 4 de JH2 (V-4JH2 en la Tabla 7)

Posición	AA progenitor (fuente)	Permitido
94	K (VH 3-23)	K
95	Y (JH2)	YSDRL (4:1:1:1:1)
96	F (JH2)	FSYDRL (4:1:1:1:1:1)
97	D (JH2)	DRSYL (4:1:1:1:1)
98	L (JH2)	LSYDR (4:1:1:1:1)
103	W (JH2)	W

15 En la Tabla 63, existe un titubeo de V-4D3-10.1a-JH2 de la Tabla 8. En 94, se permiten Lys y Arg, con Lys (el progenitor) cuatro veces tan probable como Arg. En 95, D3-10.1a (es decir, D3-10 en el primer marco de lectura y partiendo de un AA 1) tiene Leu; se permite SYDR así como Leu 4-X tan probable como cada uno de los otros tipos de AAs. En 96, D3-10.1a tiene nuevamente Leu, y se permite el mismo menú. En 97, D3-10.1a tiene Trp, y se permiten Ser, Tyr, Asp, y Arg con Trp 4-X igual de probable. En 98, D3-10.1a tiene Phe, y se permiten igualmente Ser, Tyr, Asp, y Arg.

20 Tabla 63: CDR3 de HC de longitud cuatro de V-4D3-10.1a en la Tabla 8

Posición	AA progenitor (fuente)	Permitido
94	K (VH 3-23)	KR (4:1)
95	L (D3-10.1a)	LSYDR (4:1:1:1:1)
96	L (D3-10.1a)	LSYDR (4:1:1:1:1)
97	W (D3-10.1a)	WSYDR (4:1:1:1:1)

Posición	AA progenitor (fuente)	Permitido
98	F (D3-10.1a)	FSYDR (4:1:1:1:1)
103	W	W

Ejemplo 4.4: CDR3 de HC de longitud 10 a 20

CDR3 de HC

Dos sub-bibliotecas, ambas con CDR3 de longitud 16:

5 Tabla 52: Biblioteca 1: Diversidad = 5 E 11, la secuencia "progenitora" se produce en 1 en 1,5 E 6.

Posición	AA "progenitor" (fuente)	Permitido
94	K (3-X más probable) (3-23)	KR (3:1)
95	Y (3-X más probable) (D2-21(2))	YSRDL (3:1:1:1:1)
96	Y (3-X más probable) (D2-21(2))	YSRDL (3:1:1:1:1)
97	Y (3-X más probable) (D2-21(2))	YSRDL (3:1:1:1:1)
98	D (3-X más probable) (D2-21(2))	DYSRL (3:1:1:1:1)
99	S (3-X más probable) (D2-21(2))	SYRDL (3:1:1:1:1)
100	S (3-X más probable) (D2-21(2))	SYRDL (3:1:1:1:1)
101	G (3-X más probable) (D2-21(2))	GASYRDL (3:1:1:1:1:1:1)
102	Y (3-X más probable) (D2-21(2))	YSRDL (3:1:1:1:1)
102a	Y (3-X más probable) (D2-21(2))	YSRDL (3:1:1:1:1)
102b	Y (3-X más probable) (D2-21(2))	YSRDL (3:1:1:1:1)
102c	A (3-X más probable) (JH1)	ASYRD (3:1:1:1:1)
102d	E (3-X más probable) (JH1)	ERSYL (3:1:1:1:1)
102e	Y (3-X más probable) (JH1)	YSRDL (3:1:1:1:1)
102f	F (3-X más probable) (JH1)	FYSRD (3:1:1:1:1)
102g	Q (3-X más probable) (JH1)	QERSY (3:1:1:1:1)
102h	H (3-X más probable) (JH1)	HERSYL (3:1:1:1:1:1)
103	W (JH1, fijo)	W

Tabla 53: Biblioteca 2: CDR3 de longitud 16; la diversidad es 3,0 E 10, y la secuencia progenitora se produce una vez en 3,7 E 5.

Posición	AA "progenitor" (fuente)	Permitido
94	K (3-X más probable) (3-23)	KR (3:1)
95	G (3-X más probable) (D2-2(2))	GSYDRL (3:1:1:1:1:1)
96	Y (3-X más probable) (D2-2(2))	YSDRL (3:1:1:1:1)
97	C (fijo) (D2-2(2))	C
98	S (3-X más probable) (D2-2(2))	SYRDL (3:1:1:1:1)

Posición	AA "progenitor" (fuente)	Permitido
99	S (3-X más probable) (D2-2(2))	SYRDL (3:1:1:1:1)
100	T (3-X más probable) (D2-2(2))	TYRDL (3:1:1:1:1)
101	S (3-X más probable) (D2-2(2))	SYRDL (3:1:1:1:1)
102	C (fijo) (D2-2(2))	C
102a	Y (3-X más probable) (D2-2(2))	YSDRL (3:1:1:1:1)
102b	T (3-X más probable) (D2-2(2))	TYRDL (3:1:1:1:1)
102c	A (3-X más probable) (JH1)	ASYDRL (3:1:1:1:1:1)
102d	E (3-X más probable) (JH1)	ERSYL (3:1:1:1:1)
102e	Y (3-X más probable) (JH1)	YSDRL (3:1:1:1:1)
102f	F (3-X más probable) (JH1)	FYSRDL (3:1:1:1:1:1)
102g	Q (3-X más probable) (JH1)	QERSYL (3:1:1:1:1:1)
102h	H (3-X más probable) (JH1)	HDRSYL (3:1:1:1:1:1)
103	W ((JH1))	W

La Tabla 65 muestra una variegación de titubeo de SEQ ID NO:898. La diversidad total permitida es 2,1 E 13. Una síntesis que produce 1. E 8, 3. E 8, 5. E 8, 1. E 9, o 5. E 9 muestrearán la diversidad adecuadamente. El diseño de SEQ ID NO:898 se discutió anteriormente. En el titubeo de SEQ ID NO:898, se permite el tipo de AA progenitor tres veces por encima de otros tipos de AAs en la mayoría de las posiciones. En las posiciones en las que el progenitor es Tyr, entonces se usa Tyr y Ser en cantidades iguales con Leu en la mitad de esa frecuencia. Los restos de Cys están fijos. Cada tipo de AA progenitor se permite para que haya uno de Arg, Asp, Ser, Tyr, o Leu (Leu se puede omitir si el progenitor es hidrófobo, tal como Phe). La secuencia progenitora ocurrirá una vez en 1. E 8 miembros. Las secuencias menos probables se producirán una vez en 9,5 E 16. No es importante que la biblioteca contenga realmente la secuencia progenitora, sólo que contenga muchas secuencias que se asemejen al progenitor. De este modo, una biblioteca que contiene 1. E 7, 5. E 7, 1.E8, 3. E8, 1. E 9, o 5. E 9, cuando se combina con diversidad en CDR1 de HC, CDR2 de HC, CDR1 de LC, CDR2 de LC, y CDR3 de LC, proporcionará una biblioteca que contendrá muchos Abs valiosos.

Tabla 65: Titubeo del Diseño 1 con SEO ID NO:898 como progenitora

Posición	Progenitor (fuente)	Permitido
94	K (VH 3-23)	K
95	D (No fuente)	DSYL (3:1:1:1)
96	Y (No fuente)	YSL (2:2:1)
97	G (D2-2.2)	GSYDRL (3:1:1:1:1:1)
98	Y (D2-2.2)	YSL (2:2:1)
99	C (D2-2.2)	C
100	S (D2-2.2)	SYDRL (3:1:1:1:1)
101	S (D2-2.2)	SYDRL (3:1:1:1:1)
102	T (D2-2.2)	TYDRL (3:1:1:1:1)
102a	S (D2-2.2)	SYDRL (3:1:1:1:1)
102b	C (D2-2.2)	C

Posición	Progenitor (fuente)	Permitido
102c	Y (D2-2.2)	Y SL (2:2:1)
102d	T (D2-2.2)	TY DRL (3:1:1:1:1)
102e	Y (No fuente)	Y DSL (3:1:1:1)
102f	G (No fuente)	G SYRD (3:1:1:1:1)
102g	Y (No fuente)	Y SL (2:2:1)
102h	S (No fuente)	S YDRL (3:1:1:1)
102i	Y (No fuente)	Y SL (2:2:1)
102j	A (JH1)	A SYDR (3:1:1:1:1)
102k	E (JH1)	E RSYL (3:1:1:1:1)
102l	Y (JH1)	Y SL (2:2:1)
102m	F(JH1)	F SYDR (3:1:1:1:1)
102n	Q (JH1)	Q YSDRL (3:1:1:1:1:1)
102p	H (JH1)	H SYDRL (3:1:1:1:1:1)
103	W (JH1, FR4)	W

Ejemplo 4.5: Titubeo de yycakGSGYCSCGGSCYSFDYwgqgtltvss (SEQ ID NO:931)

La Tabla 80 muestra el titubeo de SEQ ID NO:931, un ejemplo de una CDR3 de HC de longitud 15. La posición 94 es parte de FR3 y se mantiene constante. Las posiciones 95 y 96 tienen tipos de aminoácidos “progenitores” recogidos del conjunto muy usado de (YGDRS), y son G95 y S96. Las siguientes diez posiciones se toman de D2-15.2 (un segmento D usado de forma moderadamente alta que contiene un bucle cerrado por disulfuro). Las tres posiciones finales son de las posiciones 100, 101, y 102 de JH4, como se muestra en la Tabla 3. En cada posición, se hizo al tipo de aminoácido progenitor tres veces más probable que los otros tipos permitidos. Los restos de Cys están fijos. En 102e, Phe es tres veces más probable que lo son YGSRD (es decir, Phe es tres veces más probable que lo que lo son cualquiera de los aminoácidos Y, G, S, R, o D). La diversidad permitida es 1,46 E 9. La secuencia progenitora se espera en 1 en 6,9 E 4. Cada una de las secuencias individualmente sustituidas es alrededor de 1/3 igual de probable; las doblemente sustituidas son 1/9 igual de probables, y así sucesivamente. Las secuencias que están compuestas totalmente de otros tipos de AAs aparecen en solamente 1 en 1,1 E 11.

Cada una de las otras secuencias en la Tabla 21 se puede titubear digitalmente de la misma manera.

Tabla 80: Titubeo digital de yycakGSGYCSCGGSCYSFDYwgqgtltvss (SEQ ID NO:931)

Posición	Progenitor (fuente)	Permitido
94	K (VH 3-23)	K
95	G (No fuente)	G YSRD (3:1:1:1:1)
96	S (No fuente)	S GYRD (3:1:1:1:1)
97	G (D2-15.2)	G YSRD (3:1:1:1:1)
98	Y (D2-15.2)	Y GSRD (3:1:1:1:1)
99	C (D2-15.2)	C
100	S (D2-15.2)	S GYRD (3:1:1:1:1)
101	G (D2-15.2)	G YSRD (3:1:1:1:1)
102	G (D2-15.2)	G YSRD (3:1:1:1:1)

Posición	Progenitor (fuente)	Permitido
102a	S (D2-15.2)	SGYRD (3:1:1:1:1)
102b	C (D2-15.2)	C
102c	Y (D2-15.2)	YGSRD (3:1:1:1:1)
102d	S (D2-15.2)	SGYRD (3:1:1:1:1)
102e	F (JH4)	FYGSRD (3:1:1:1:1:1)
102f	D (JH4)	DGSRY (3:1:1:1:1)
102g	Y (JH4)	YGSRD (3:1:1:1:1)
103	W (JH4, FR4)	W

Ejemplo 5: Diversidad de cadena ligera sintética

Para obtener anticuerpos completos, se necesita combinar una biblioteca de cadenas pesadas con una biblioteca de cadenas ligeras (LC). En los Abs naturales, a menudo se observa que HC lleva a cabo la mayor parte de la unión, y muchas bibliotecas han prestado poca atención a la LC o han obtenido diversidad de LC a partir de donantes humanos. Para proporcionar suficiente diversidad para dar buenos ligadores a casi cualquier diana, se ha diseñado un programa de diversificación que supera lo que proporciona habitualmente el sistema inmunitario humano. No obstante, el programa está diseñado para producir LCs completamente funcionales que tienen el mismo tipo de cambios que el observado en Abs naturales. Como la LC, se recogió sólo unos pocos más V kappa III A27.

A partir de una biblioteca que comprende LCs kappa y lambda donadas, se tipó una colección de 1266 Abs. Entre VKills. A27 se observa mayoritariamente (Tabla 66) y se empareja bien con HC 3-23.

Las CDRs de A27 contienen 12, 7, y 9 aminoácidos. La colocación de la diversidad en todas estas posiciones puede no funcionar bien: a) puede haber muchos miembros inestables o no funcionales, y b) la diversidad en algunas posiciones puede no ayudar a mejorar la unión. Se ha reducido el número de posiciones variables de 28 a 16.

Se ha estudiado la estructura tridimensional de 1 QLR que tiene una LC A27. La estructura de 1 GLR está públicamente disponible en la RCDB Protein Data Base. A partir de esto, los restos marcados en la Tabla 68 parecen útiles para ser variados. El T56 es alrededor de 10 Å desde una His en CDR3 de HC. La variación en 56 puede ser útil. G24 tiene sólo alrededor de 7 Å desde un átomo en CDR3 de HC. La línea germinal es R24; de este modo, la variación en 24 puede ser útil.

La Tabla 69 muestra un casete de presentación que se diseñó para uso en pMID21. De este modo, las enzimas de restricción recogidas no tienen otros sitios en pMID21. *SpeI* está en la secuencia señal *iii* y permite que toda la LC se inserte o sea eliminada. *XmaI*, *PpuMI*, *EcoO109I*, y *BplI* preceden a CDR1. *SacII* está en FR2, separando CDR1 de CDR2. Como alternativa, se podría insertar un sitio *AvrII* en la misma posición. Los sitios *BspEI* y *XhoI* están en FR3, y en FR4 hay un sitio *KpnI*.

Se reunieron 155 secuencias A27 y se analizó qué pasa en las CDRs. La Tabla 70 muestra el análisis. En la Tabla 70, se muestra lo que se encuentra en los Abs de nuestra biblioteca y lo que pondríamos en cada posición.

!Tabla 68: en la que A27 varía

```

:
!      22    3    3    5    5    89    9
!      45    0a  4    0    5    90    5
!1QLR      GASQSVS_NYLA  DASSRAT  QQYGSSPLT
!A27       RASQSVSSSYLA  GASSRAT  QQYGSSPLT
          ** * * * * *  * * * *
GASQSVS es (SEQID NO:922)  DASSRAT es (SEQID NO:923)
QQYGSSPLY es (SEQID NO:924)
RASQSVSSSYLA es (SEQ ID NO:925)  GASSRAT es (SEQ ID NO:926)

```

La Tabla 68 muestra dónde se varregarían las CDRs de A27.

Tabla 67: Comparación de secuencias de AA de VKIII

VK3 es (SEQ ID NO:927)

```

FR1.....CDR1.....FR2.....CDR2...FR3.....CDR3.....FR4.....
      1         2         3         4         5         6         7         8         9         10
VK3  123456789012345678901234567890123456789012345678901234567890123456789
A27  DIVLTQSPATLSLSPGERATLSCRASGVSSSYLAWYQKPGQAPRLLIYGASSRATGVFARFSGSGSGTDFTLTISSSLEPEDEFAVYYCQQHYTTPTFGCGTKVEIKRT
A11  E-----G-----G-----L-----D-----I-D-----R-----YGSS-
L2   E--M-----V-----IN-----T-----I-----E-----QS-----YNNW-
L16  E--M-----V-----IN-----T-----I-----E-----QS-----YNNW-
L6   E-----G-----D--N--I-----E-----QS-----RSNW-
L20  E-----G-----D--N--I-----P-----RSNW-
L25  E--M-----S-----T-----I-----Q-----D-NL-

```

VK3 difiere de A27 en E1D, G9A, I58V, D60A, R77S.

A27 y L6 difieren en G9A(FR1), Δ 31a (en CDR1), G50D(CDR2), S53N(CDR2), G92S(CDR3), S93N(CDR3), S94W(CDR3) VK3 del documento US 7.264.963.

CDR1

R24, A25 y S26 están demasiado lejos del sitio de combinación para ayudar, y se mantuvieron constantes. El grupo lateral de V29 está escondido; esta posición se mantuvo constante como Val. En las otras posiciones, se permitió Y o S y un movimiento de intercambio de cargas (RE o RD, dependiendo de si la muestra tuvo más de E o D en la posición en cuestión) más otros tipos que se observaron frecuentemente. Se usó una hoja de cálculo de Excel para determinar que este patrón de variegación daría la secuencia progenitora a 0,8% si los "otros" AAs se sustituyen a 5%, a 0,1% si los "otros" AAs se sustituyeron a 6,5%, y a 0,02% si los "otros" fueron a 9%. En la muestra de 155, 17 tienen un AA suprimido (incluyendo 1 QLR); de este modo, nos organizaremos para suprimir S30a en ~8% de los miembros.

CDR2

De la inspección de 1 QLR, se observa que CDR2 está en cierto modo alejada del sitio de combinación. Ha habido sugerencias de que se mantienen constantes los restos en esta CDR. El estudio de la estructura tridimensional sugiere que la variegación en G50, S53, y T56 podría ser útil. S53 es el más variable en la muestra de 155, pero esto no demuestra que estos cambios son útiles. En 1QLR, G50 se ha mutado a R50. El grupo lateral de T56 apunta hacia CDR3 de HC y está a alrededor de 11 Å de un átomo en CDR3 de HC.

CDR3

Q89 y Q90 están escondidos y la naturaleza no los varía a menudo; estos restos no varían. Y91 está empaquetado contra CDR3 de HC, y los cambios aquí alterarían el sitio de combinación y sí se producen. En G92, $\phi = -80$ y $\psi = -15$, de manera que la colocación en un no Gly es factible; la naturaleza lo hace en 47/155 casos. S93 varía o se suprime muy a menudo. S94 está muy expuesto y está muy variado. P95 está expuesto y variado, L96 se empaqueta contra CDR3 de HC: los cambios aquí afectarán al sitio de unión y sí se producen en la naturaleza. T97 está escondido y se ha mantenido constante/el aminoácido no varía.

La secuencia progenitora aparece en 0,000246 o 1 en 4,06 E 3. La diversidad permitida es alrededor de 2,1 E 12. Con dos supresiones del 8%, 84,6% de los miembros serán de longitud completa, 7,4% tendrán CDR1 corta y CDR3 de longitud completa, 7,4% tendrán CDR1 de longitud completa y CDR3 corta, y 0,6% tendrán ambas supresiones.

Otras líneas germinales no estaban en la muestra.

Tabla 66: Distribución de VLs en 1266 LCs seleccionadas.							
Kappas				Lambdas			
O12	VKI	313		1a	VL1	9	
O18	VKI	1		1e	VL1	7	
A20	VKI	26		1c	VL1	55	
A30	VKI	26		1g	VL1	46	
L14	VKI	2		1b	VL1	1	118
L1	VKI	5		2c	VL2	18	
L15	VKI	1		2e	VL2	23	

L5	VKI	83		2a2	VL2	79	
L8	VKI	10		2d	VL2	1	121
L12	VKI	77	544	3r	VL3	56	
O11	VKII	4		3j	VL3	4	
A17	VKII	17		3i	VL3	31	
A19	VKII	31	52	3h	VL3	22	113
A27	VKIII	155		4a	VL4	1	1
L2	VKIII	31		5c	VL5	1	1
L6	VKIII	88		6a	VL6	8	8
L25	VKIII	16	290	10a	VL10	6	6
B3	VKIV	12	12	Número of lambdas			368
Número of kappas			898	Abs totales en la muestra			1266

!Tabla 69: Un gen de presentación para A27 en pM21J. ! IIIsignal::A27::Ckappa

La secuencia de aminoácidos de la Tabla 69 es (SEQ ID NO:928).

La secuencia de ADN de la Tabla 69 es (SEQ ID NO:929).

```

!          secuencia señal -----
!      1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
!      M   K   K   L   L   S   A   I   P   L   V   V   P   F   Y
!
!  1 |atg|aaG|aaA|ctg|ctg|tct|gct|atc|ccA|CTA|GTt|gtc|cct|ttc|tat|
!                                     SpeI...
!
!
!      Señal----- FR1-----
!     16  17  18  19  20  21  22  23  24  25  26  27  28  29  30
!     S   H   S   E1  I   V3  L   T5  Q   S7  P   G9  T   L   S12
!
!  46 |tct|cat|agt|gaa|atc|ggt|ctg|acc|cag|tcC|CCG|GGG|aCC|Ctg|tgt|
!                                           XmaI...
!                                           PpuMI...
!                                           EcoO109I. (1/2)

```

La secuencia de aminoácidos de la Tabla 69 es (SEQ ID NO:928).

La secuencia de ADN de la Tabla 69 es (SEQ ID NO:929).

```

!      FR1----- CDR1-----
!      31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
!      L13 S  P  G  E  R  A  T  L  S  C23 R24 A  S  Q
!  91  |ctg|tct|ccg|ggg|gaa|cgt|gct|acG|CTg|AGC|tgt|cgt|gct|tct|caa|
!      B1pI....
!
!      CDR1----- FR2-----
!      46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
!      S28 V  S  S  S30a Y  L  A34 W  Y  Q  Q  K  P  G
!  136 |tcc|gtt|agC|TCC|TCt|tat|tta|gct|tgg|tat|cag|caa|aag|ccg|ggg|
!      BseRI...
!
!      FR2----- CDR2-----
!      61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
!      Q  A  P  R45 L  L  I  Y  G50 A  S  S  R  A  T56
!  181 |caa|gct|CCG|CGG|ctg|ttg|atc|tat|ggg|gcc|tct|agt|cgt|gct|act|
!      SacII..
!
!      FR3-----
!      76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
!      G  I  P  D60 R  F  S  G  S65 G  S  G  T  D  F
!  226 |ggc|atc|cct|gat|cgt|ttc|tct|ggc|tct|ggc|TCC|GGA|acc|gat|ttc|
!      BspEI..
!
!      FR3-----
!      91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
!      T  L  T  I  S  R  L  E  P  E  D  F  A  V  Y
!  271 |act|ctg|acc|att|tct|CGT|CTC|GAG|ccg|gaa|gat|ttc|gct|gtc|tac|
!      BsmBI..
!      XhoI...
!
!      FR3----- CDR3----- FR4-----
!      106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
!      Y  C  Q89 Q  Y  G  S  S  P95 L  T  F  G  G  G
!  316 |tat|tgt|caa|cag|tat|ggg|tct|agt|ccg|ctg|act|ttc|ggg|ggc|GGT|
!      KpnI...
!
!      FR4-----
!      121 122 123 124 125 126
!      T  K  V  E  I  K
!  361 |ACC|aaa|gtc|gaa|atc|aag
!      KpnI.
!
!      Ckappa-----
!      R  G  T  V  A  A  P  S  V  F  I  F  P  P  S
!  379 cgt gga act gtg gCT GCA Cca tct GTC TTC atc ttc ccg cca tct
!      BsgI.... BbsI....
!
!      D  E  Q  L  K  S  G  T  A  S  V  V  C  L  L
!  424 gat gag cag ttg aaa tct gga act gcc tct gtt gtg tgc ctg ctg
!
!      N  N  F  Y  P  R  E  A  K  V  Q  W  K  V  D
!  469 aat aac ttc tat ccc aga gag gcc aaa gta cag tgg aag gtg gat
!
!      N  A  L  Q  S  G  N  S  Q  E  S  V  T  E  Q
!  514 aac gcc ctc caa tcg ggt aac tcc cag gag agt gtc aca gag cag

```


La secuencia de aminoácidos de la Tabla 69 es (SEQ ID NO:928).

La secuencia de ADN de la Tabla 69 es (SEQ ID NO:929).

```

!      D   S   K   D   S   T   Y   S   L   S   S   T   L   T   L
559    gac agc aag gac agc acc tac agc ctc agc agc acc ctg act ctg
!
!      S   K   A   D   Y   E   K   H   K   V   Y   A   C   E   V
604    tcc aaa gca gac tac gag aaa cac aaa GTC TAC gcc tgc gaa gtc
!
!      T   H   Q   G   L   S   S   P   V   T   K   S   F   N   R
649    acc cat cAG GGC CTg agt tCA CCG GTG aca aag agc ttc aac agg
!      AlwNI..... SgrAI.....
!      EcoO109I.(2/2) AgeI.....
!
!      G   E   C   .   .
694    gga gag tgt taa taa
!
!      GG CGCGCCaatt
709                                AscI.....
!                                BssHII.
!

```

Tabla 70: Cómputo de mutaciones en CDRs de Abs A27

CDR1

R24	1,	3G, 1T, 151-,	Fix
A25	2,	3T, 152-,	Fix
S26	3,	1R, 154-,	Fix
Q27	4,	3E , 1H, 1L, 1P, 4R , 145-,	9% ERYSL
S28	5,	1A, 2F, 2G, 1I, 2L, 5N , 1P, 1R, 10T , 1V, 1Y , 128-,	9% NTYERL
V29	6,	1F, 19I, 6L, 129-,	Fix
S30	7,	2A, 2D , 8G, 2H, 1I, 11N , 9R , 6T , 4V , 2Y , 108 -,	9% DNRTY
S30a	8,	1A, 2F, 6G , 1H, 6N , 1P, 10R , 6T , 3Y , 119 -,	9% GNRTYD
(8% suprime 30a)			
S31	9,	1A, 5D , 3F , 4G , 1H, 2I, 4K, 1L, 31N , 19R , 7T , 7Y , 70 -,	9% DFGNRTY
Y32	10,	5F, 1K, 14L, 4N, 4Q, 2R, 8S, 3V, 1W, 113 -,	9% FDLNQRSY
L33	11,	16A , 1F, 4I, 1N, 1S, 8V, 1Y, 123 -,	Fix
A34	12,	2G, 2L, 1N, 1S, 4V, 128-,	9% SY
-	13,	2A, 1G,	
-	14,	1S,	
-	15,	1S,	
-	16,	1Y,	
-	17,	1L,	
-	18,	1A,	

Nota: un anticuerpo tuvo una inserción de seis AAs en CDR1! Otros dos Abs tuvieron una única inserción. Diecisiete Abs tuvieron una supresión de AA en CDR1.

CDR2

ES 2 528 963 T3

G50	1,	10A, 11D, 1H, 2R, 2S, 1V, 7Y, 121- ,	9%	DRSYL
A51	2,	7G, 2I, 6S, 7T, 2V, 131-,		Fix
S52	3,	6A, 3F, 1G, 1T, 144-,		Fix
S53	4,	1A, 1G, 1H, 5I, 2K, 16N, 7R, 16T, 106- ,	9%	NTSYER
R54	5,	1A, 1I, 1N, 1S, 3T, 1Y, 147-,		Fix
A55	6,	2P,7R,4S,2V,140-,		Fix
T56	7,	10A ¹ , 1G, 1H, 2P, 4S, 137-,	9%	ERSY
-	8,	1A, 6T,		

Nota: hay siete anticuerpos con una inserción de un AA.

CDR3 (al mostrar “_” se quiere decir que el Ab tiene una supresión en CDR3)

Q89	1,	5H, 1L, 2M, 147-,		Fix
Q90	2,	1E, 1F, 13H, 2K, 2L, 4R, 1S, 1Y, 130-,		Fix
Y91	3,	2A, 8F, 2G, 2H, 1L, 1P, 13R, 4S, 122- ,	9%	FERS
G92	4,	10A, 3D, 2H, 1I, 1L, 2N, 6R, 12S, 2V, 3Y, 108- , 5_,	9%	ADRSTY
S93	5,	1A, 2D, 2F , 6G, 2H, 3I, 2K, 2M, 14N , 1P, 1Q, 8R, 17T, 2Y, 86- , 6_,	9%	DFNRTY
(8% tienen 93 suprimido)				
S94	6,	3A, 6F, 1I, 3L, 3P, 2R, 2T, 11W, 117- , 7_,	9%	WERYS
P95	7,	2A, 1E, 1G, 1K, 3L, 1M, 7R, 3S, 3T, 1V, 108- , 24_,	9%	ERYSL
L96	8,	2A, 2E , 3F, 3G, 1H, 1I, 3K , 7L, 2M, 24P² , 6Q, 28R , 3S, 3T, 7V, 2Y, 58- ,	9%	ERPYS
T97	9,	2A, 1F, 2G, 3I, 1K, 2L, 3M, 1N, 1R, 6S, 3V, 2Y, 128- ,		Fix
-	10,	1A, 1S, 34-,		
-	11,	1S, 7-,		
-	12,	1A,		

Hay un Ab con una inserción de 3 AAs.

Cinco tienen supresiones de 4 AAs, 1 tiene una supresión de 3 AAs, 1 tiene una supresión de 2 AAs, y 17 tienen una supresión de un AA.

Tabla 71: Diversidad permitida en CDR1, 2, y 3 de A27::JK4.

Posición	pprogenitor	permitido	
CDR1			
42(24)	R	fijo	
43 (25)	A	fijo	
44 (26)	S	fijo	
45 (27)	Q	ERYSL	55%Q 9% otros
46 (28)	S	NTYERL	46%S 9% otros
47 (29)	V	fijo	
48 (30)	S	DNRTY	55%S 9% otros

49 (30a)	S	GNRTYD	46%S 9% otros
8% han suprimido 30a			
50 (31)	S	DFGNRTY	44%S 8% otros
51 (32)	Y	FDLNQRSY	44%Y 7% otros
52 (33)	L	fijo	
53 (34)	A	SY	70%A 15% otros
CDR2			
69 (50)	G	DRSYL	55%G 9% otros
70 (51)	A	Fijo	
71 (52)	S	Fijo	
72 (53)	S	NTSYER	52%S 8% otros
73 (54)	R	Fijo	
74 (55)	A	Fijo	
75 (56)	T	ERSY	64%T 9% otros
CDR3			
108 (89)	Q	fijo	
109 (90)	Q	fijo	
110 (91)	Y	FERS	64%Y 9% otros
111 (92)	G	ADRSTY	52%G 8% otros
112 (93)	S	DFNRTY	52%S 8% otros
113 (94)	S	WERYS	55%S 9% otros
114 (95)	P	ERYS	64%P 9% otros
8% han suprimido P95			
115 (96)	L	ERPYS	55%L 9% other
116 (97)	T	fijo	
¹ Siete de estos proceden de inserciones.			
² Algunos de estos aparecen debido a inserciones.			

La secuencia progenitora aparece en $5,32 \times 10^{-5}$ ó 1 en $1,88 \times 10^4$.

Las secuencias con una única sustitución tienen una probabilidad entre $1,1 \times 10^{-5}$ y $7,5 \times 10^{-6}$.

Las secuencias que no tienen ninguno de los AAs progenitores aparecen en 1 en $6,7 \times 10^6$.

5 La diversidad permitida es alrededor de $2,35 \times 10^2$.

Tabla 75: Frecuencias de aminoácidos en CDR3s de HC

AA	Número	%	Rel up	Rel down
Y	3428	15,64	50,41	1,00

ES 2 528 963 T3

G	3244	14,80	47,71	0,95
D	2622	11,96	38,56	0,76
S	1777	8,11	26,13	0,52
R	1337	6,10	19,66	0,39
F	1328	6,06	19,53	0,39
A	1213	5,53	17,84	0,35
V	1141	5,20	16,78	0,33
L	816	3,72	12,00	0,24
I	745	3,40	10,96	0,22
P	726	3,31	10,68	0,21
T	586	2,67	8,62	0,17
W	566	2,58	8,32	0,17
M	560	2,55	8,24	0,16
N	462	2,11	6,79	0,13
E	363	1,66	5,34	0,11
K	355	1,62	5,22	0,10
H	327	1,49	4,81	0,10
Q	259	1,18	3,81	0,08
C	68	0,31	1,00	0,02
Total	21923			

Tabla 76: Distribución de longitudes de CDR3 de HC

Longitud	Número de anticuerpos	Suma	Mediana
1	0		
2	0		
3	2	2	
4	21	23	
5	16	39	
6	100	139	
7	36	175	
8	78	253	
9	155	408	
10	153	561	
11	134	695	11,12
12	123	818	

ES 2 528 963 T3

13	133	951	
14	92	1043	
15	87	1130	
16	71	1201	
17	59	1260	
18	41	1301	
19	40	1341	
20	22	1363	
21	21	1384	
22	15	1399	
23	7	1406	
24	7	1413	
25	1	1414	
26	1	1415	
27	3	1418	
28	0	1418	
29	0	1418	
30	1	1419	
	1419	709,5	

Tabla 77: Utilización de segmentos D (corte a 0,70 de coincidencia)

3-22.2	38	YYYDSSGYYY
4-17.2	27	DYGDY
3-3.2	25	YYDFWSGYYT
6-19.1	25	GYSSGWY
7-27.1	19	LTG
5-5.3	18	GYSYGY
6-13.1	18	GYSSSWY
5-12.3	13	GYSGYDY
6-13.2	10	GIAAAG
1-26.3	9	YSGSY
2-15.2	9	GYCSGGSCYS
4-4.3	9	TTVT
3-10.2	8	YYYGSGSYN
1-1.3	7	YNWND

4-4.2	7	DYSNY
2-2.2	6	GYCSSTSCYT
3-16.2	6	YYDYVWGSYRYT
6-6.1	6	EYSSSS
6-19.2	6	GIAVAG
3-9.1	5	VLRYFDWLL@
4-23.2	5	DYGGNS
6-6.2	5	SIAAR
1-7.3	4	YNWNY
2-2.3	4	DIWVPAAI
4-23.3	4	TTVVT
1-7.1	3	GITGT
1-26.1	3	GIVGAT
7-27.3	3	NWG
3-10.1	2	VLLWFGELL@
3-10.2	2	ITMVRGVII
5-5.1	2	VDTAMV
5-5.2	2	WIQLWL
5-12.1	2	VDIVATI
5-24.3	2	RDGYNY
1-1.1	1	GTTGT
2-21.3	1	HIVVTAI
3-3.3	1	ITIFGVVII
5-24.2	1	*RWLQL
6-6.3	1	V*QLV
6-19.3	1	V*QWLV

Tabla 78: Utilización de segmentos D (corte a 0,667)

Nombre	Número	Secuencia	%
Ninguno	935		0,517
7-27.1	158	LTG	0,087
7-27.3	98	NWG	0,054
5-5.3	72	GYSYGY	0,040
1-26.3	67	YSGSYY	0,037
3-22.2	46	YYYDSSGYYY	0,025

ES 2 528 963 T3

4-17.2	38	DYGDY	0,021
3-3.2	37	YYDFWSGYT	0,020
7-27.2	37	@ LG	0,020
6-19.1	33	GYSSGWY	0,018
6-13.2	31	GIAAAG	0,017
6-13.1	22	GYSSSWY	0,012
6-6.1	18	EYSSSS	0,010
6-19.2	18	GIAVAG	0,010
4-23.2	17	DYGGNS	0,009
5-12.3	17	GYSGYDY	0,009
5-24.3	14	RDGYNY	0,008
2-15.2	13	GYCSGGSCYS	0,007
1-26.1	11	GIVGAT	0,006
4-4.3	11	TTVT	0,006
1-1.3	9	YNWND	0,005
2-2.2	9	GYCSSTSCYT	0,005
3-16.2	9	YYDYVWGSYRYT	0,005
2-2.3	8	DIVVPAAI	0,004
3-10.2	8	YYYGSGSYYN	0,004
4-4.2	8	DYSNY	0,004
1-7.3	7	YNWNY	0,004
3-3.3	6	ITIFGVVII	0,003
6-6.2	6	SIAAR	0,003
3-9.1	5	VLRYFDWLL@	0,003
3-10.2	5	ITMVRGVII	0,003
6-19.3	5	V*QWLV	0,003
1-7.1	4	GITGT	0,002
4-23.3	4	TTVVT	0,002
1-1.1	3	GTTGT	0,002
5-5.1	3	VDTAMV	0,002
5-24.2	3	*RWLQL	0,002
3-10.1	2	VLLWFGELL@	0,001
5-5.2	2	WIQLWL	0,001
5-12.1	2	VDIVATI	0,001
1-26.2	1	V*WELL	0,001

2-21.2	1	AYCGGDCYS	0,001
2-21.3	1	HVVVTAI	0,001
3-3.1	1	VLRFFLEWLLY	0,001
3-16.2	1	IMITFGGVIVI	0,001
6-6.3	1	V*QLV	0,001
6-13.3	1	V*QLLV	0,001

Tabla 78: Utilización de segmentos JH

JH1	17
JH2	31
JH3	452
JH4	636
JH5	32
JH6	251

Ejemplo 6: ADN titubeado para CDR3 de HC 16d

La Tabla 400 muestra un segmento de ADN de un sitio *Xba*I en FR3 a un sitio *Bst*ElI en FR4. La CDR3 de HC consiste en SYSY::D2-2(2)::QH seguido de la región FR4 de JH1. El QH se encuentra en la línea germinal de JH1. En la unión de V-D-J, las células inmunitarias a menudo editan los extremos de V, D, y J. De este modo, la construcción corresponde a lo que es muy posible en la construcción y maduración de genes de inmunoglobulina real. Titubeando la síntesis, se obtiene una gran colección de genes que se asemejan a lo que provendría de unir 3-23 a una región D o a un JH1 poco editado seguido de algunas mutaciones. En la biblioteca 16d, hay dos cisteínas que forman presuntamente un disulfuro; éstas no están titubeadas.

La Tabla 500 muestra la distribución esperada de tipos de aminoácidos en cada posición en la biblioteca 16d. El dopaje del titubeo se ajustó a 73:9:9:9. La secuencia más probable es la mostrada en la Tabla 21, y debería estar presente a una frecuencia de 4,8 E -5. Sólo 55% de las secuencias están libres e parada, y 74% están libres de ochre u opel. Si la biblioteca se expresa en células supE, esto es un número importante. Sería valioso eliminar las secuencias con codones de parada como se discute en otra parte aquí. Se puede observar que esas posiciones que comienzan como S se predice que tendrán S 54% de las veces e Y 5,4%, mientras que aquellas que comienzan como Y tienen Y 44% de las veces y S 7,2%. En cada posición hay 7-9 tipos de AAs que aparecen a >1%. Hay 14 posiciones variegadas. Las secuencias que se muestrearán más eficazmente ascienden a alrededor de $8^{14} = 4,3 \text{ E } 12$.

Tabla 400: Casete para presentación de CDR3 de HC titubeada 16d

```

!      -----FR3-----
!      68 69 70 71 72 73 74 75 76 77 78 79 80 81 82
!      T  I  S  R  D  N  S  K  N  T  L  Y  L  Q  M
1216 |act|atc|TCT|AGA|gac|aac|tct|aag|aat|act|ctc|tac|ttg|cag|atg|
!      | xbaI |
!
!      -----FR3----->|
!      82a 82b 82c 83 84 85 86 87 88 89 90 91 92 93 94
!      N  S  L  R  A  E  D  T  A  V  Y  Y  C  A  K
1261 |aac|agC|TTA|AGg|gct|gag|gac|act|gca|gtc|tac|tat|tgc|gct|aaa|
!      |AflIII |
!

```



```

e = 0,73 A + 0,09 C + 0,09 G + 0,09 T
q = 0,09 A + 0,73 C + 0,09 G + 0,09 T
j = 0,09 A + 0,09 C + 0,73 G + 0,09 T
z = 0,09 A + 0,09 C + 0,09 G + 0,73 T
Los valores 0,73 y 0,09 se recogen de manera que 0,73 + 3*0,09 = 1,0
Se podrían usar otras relaciones

102 102 102 102 102 102 102 102
95 96 97 98 99 100 101 102 a b c d e f g h
S Y S Y G Y c S S T S c Y T Q H
zqz zez zqz zez jjz zez TGT zqz zqz eqz zqz TGT zez eqz qej qez

-----FR4----->|
103 104 105 106 107 108 109 110 111 112 113
W G Q G T L V T V S S
|TGg|ggg|caa|ggg|act|ttG|GTC|ACC|gtc|tct|agt
| BstEII |

```

Tabla 500: Distribución esperada de tipos de AAs en CDR3 de HC titubeada

"." = TGA o TAA; "b" = TAG

S Y S Y G Y c S S T S c Y T Q H
zqz zez zqz zez jjz zez tgt zqz zqz eqz zqz tgt zez eqz qej qez

Pureza de base nominal = 0.7300 Otros = 0.0900

s (zqz)	y (zez)	s (zqz)	y (zez)	g (jjz)	y (zez)	C (TGT)	s (zqz)	s (zqz)	t (eqz)
1 s 5.4-01	y 4.4-01	s 5.4-01	y 4.4-01	g 5.3-01	y 4.4-01	c 1.000	s 5.4-01	s 5.4-01	t 5.3-01
2 p 6.6-02	s 7.2-02	p 6.6-02	s 7.2-02	r 7.8-02	s 7.2-02		p 6.6-02	p 6.6-02	s 1.2-01
3 a 6.6-02	f 5.4-02	a 6.6-02	f 5.4-02	a 6.6-02	f 5.4-02		a 6.6-02	a 6.6-02	a 6.6-02
4 t 6.6-02	h 5.4-02	t 6.6-02	h 5.4-02	v 6.6-02	h 5.4-02		t 6.6-02	t 6.6-02	p 6.6-02
5 f 5.4-02	n 5.4-02	f 5.4-02	n 5.4-02	s 6.2-02	n 5.4-02		f 5.4-02	f 5.4-02	i 6.0-02
6 c 5.4-02	c 5.4-02	c 5.4-02	c 5.4-02	c 5.4-02	c 5.4-02		c 5.4-02	c 5.4-02	n 5.4-02
7 y 5.4-02	d 5.4-02	y 5.4-02	d 5.4-02	d 5.4-02	d 5.4-02		y 5.4-02	y 5.4-02	r 2.0-02
8 l 2.0-02	. 5.4-02	l 2.0-02	. 5.4-02	e 1.2-02	. 5.4-02		l 2.0-02	l 2.0-02	k 1.2-02
9 . 1.2-02	b 4.8-02	. 1.2-02	b 4.8-02	l 9.6-03	b 4.8-02		. 1.2-02	. 1.2-02	l 9.6-03
10 r 9.6-03	l 2.0-02	r 9.6-03	l 2.0-02	t 8.1-03	l 2.0-02		r 9.6-03	r 9.6-03	g 8.1-03
11 g 8.1-03	k 1.2-02	g 8.1-03	k 1.2-02	p 8.1-03	k 1.2-02		g 8.1-03	g 8.1-03	v 8.1-03
12 v 8.1-03	q 1.2-02	v 8.1-03	q 1.2-02	i 7.4-03	q 1.2-02		v 8.1-03	v 8.1-03	f 6.6-03
13 i 7.4-03	e 1.2-02	i 7.4-03	e 1.2-02	. 6.6-03	e 1.2-02		i 7.4-03	i 7.4-03	c 6.6-03
14 h 6.6-03	r 9.6-03	h 6.6-03	r 9.6-03	f 6.6-03	r 9.6-03		h 6.6-03	h 6.6-03	h 6.6-03
15 n 6.6-03	t 8.1-03	n 6.6-03	t 8.1-03	h 6.6-03	t 8.1-03		n 6.6-03	n 6.6-03	d 6.6-03
16 d 6.6-03	v 8.1-03	d 6.6-03	v 8.1-03	y 6.6-03	v 8.1-03		d 6.6-03	d 6.6-03	y 6.6-03
17 w 5.9-03	a 8.1-03	w 5.9-03	a 8.1-03	n 6.6-03	a 8.1-03		w 5.9-03	w 5.9-03	m 5.9-03
18 b 5.9-03	g 8.1-03	b 5.9-03	g 8.1-03	w 5.9-03	g 8.1-03		b 5.9-03	b 5.9-03	q 1.5-03
19 q 1.5-03	p 8.1-03	q 1.5-03	p 8.1-03	q 1.5-03	p 8.1-03		q 1.5-03	q 1.5-03	e 1.5-03
20 k 1.5-03	i 7.4-03	k 1.5-03	i 7.4-03	k 1.5-03	i 7.4-03		k 1.5-03	k 1.5-03	. 1.5-03
21 e 1.5-03	w 5.9-03	e 1.5-03	w 5.9-03	m 7.3-04	w 5.9-03		e 1.5-03	e 1.5-03	w 7.3-04
22 m 7.3-04	m 7.3-04	m 7.3-04	m 7.3-04	b 7.3-04	m 7.3-04		m 7.3-04	m 7.3-04	b 7.3-04

ES 2 528 963 T3

s (zqz)	C (TGT)	y (zez)	t (eqz)	q (qej)	h (qez)
1 s 5.4-01	c 1.000	y 4.4-01	t 5.3-01	q 4.4-01	h 4.4-01
2 p 6.6-02		s 7.2-02	s 1.2-01	h 9.6-02	q 9.6-02
3 a 6.6-02		f 5.4-02	a 6.6-02	l 7.2-02	l 6.7-02
4 t 6.6-02		h 5.4-02	p 6.6-02	r 7.2-02	r 6.7-02
5 f 5.4-02		n 5.4-02	i 6.0-02	p 6.6-02	p 6.6-02
6 c 5.4-02		c 5.4-02	n 5.4-02	e 5.4-02	n 5.4-02
7 y 5.4-02		d 5.4-02	r 2.0-02	k 5.4-02	d 5.4-02
8 l 2.0-02		. 5.4-02	k 1.2-02	b 4.8-02	y 5.4-02
9 . 1.2-02		b 4.8-02	l 9.6-03	d 1.2-02	s 1.5-02
10 r 9.6-03		l 2.0-02	g 8.1-03	y 1.2-02	k 1.2-02
11 g 8.1-03		k 1.2-02	v 8.1-03	n 1.2-02	e 1.2-02
12 v 8.1-03		q 1.2-02	f 6.6-03	s 9.6-03	g 8.1-03
13 i 7.4-03		e 1.2-02	c 6.6-03	t 8.1-03	t 8.1-03
14 h 6.6-03		r 9.6-03	h 6.6-03	v 8.1-03	v 8.1-03
15 n 6.6-03		t 8.1-03	d 6.6-03	a 8.1-03	a 8.1-03
16 d 6.6-03		v 8.1-03	y 6.6-03	g 8.1-03	i 7.4-03
17 w 5.9-03		a 8.1-03	m 5.9-03	. 6.6-03	. 6.6-03
18 b 5.9-03		g 8.1-03	q 1.5-03	w 5.9-03	c 6.6-03
19 q 1.5-03		p 8.1-03	e 1.5-03	m 5.9-03	f 6.6-03
20 k 1.5-03		i 7.4-03	. 1.5-03	i 2.2-03	b 5.9-03
21 e 1.5-03		w 5.9-03	w 7.3-04	f 1.5-03	w 7.3-04
22 m 7.3-04		m 7.3-04	b 7.3-04	c 1.5-03	m 7.3-04

La secuencia más probable tiene frecuencia = 4.8E-05
 Libre de parada de fracción = 5.5E-01
 Libre de (TAA y TGA) de fracción = 7.4E-01

	F%	F%	F%
D1 1-1	0.42	0.14	2.90
1-7	0.42	0.28	1.24
1-20	0.00	0.00	0.00
1-26	0.00	0.97	1.80
D2 2-2	0.55	4.30	1.21
2-8	0.00	0.67	0.41
2-15	0.28	4.03	0.94
2-21	0.00	2.22	0.94
D3 3-3	0.94	4.44	3.70
3-9	0.67	1.82	0.00
3-10	0.67	5.78	1.55
3-16	1.08	2.49	0.67
3-22	0.14	7.87	0.81
D4 4-4	0.28	0.69	0.28
4-11	0.00	0.00	0.00
4-17	0.00	4.03	2.76
4-23	0.14	1.41	0.54
D5 5-5	1.34	0.40	4.30
5-12	1.08	0.00	1.95
5-18	0.00	0.00	0.00
5-24	0.67	1.55	1.82
D6 6-6	1.21	1.55	0.13
6-13	4.84	2.62	0.27
6-19	6.66	1.95	0.54
D7 7-27	0.27	0.13	0.27
Total fraccional %	21.65	49.34	29.01

!Tabla 800: LC K1(O12)::JK1

```

!      ... Sec. líder  ->|----- FR1 ----->
!      1 2 3 4 5 6 7 8 9 10 11
!      G V H S A Q D I Q M T Q S P S S L
1  |ggT|GTA|CAc|aGT|GCT|Cag|gat|att|cag|atg|act|caa|tct|ccC|TCG|AGt|ctg|
   |BsrGI...   ApaLI...   XhoI....
!
!      ----- FR1 ----->|--- CDR1 --->
!      12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
!      S A S V G D R V T I T C R A S
46 |tct|gct|tct|gtc|gGC|GAT|CGC|gtt|act|att|act|tgt|cgt|gct|tcc|
   |SgfI.....
!
!      ---- CDR1 ----->|---- FR2 ----->
!      27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
!      Q S I S S Y L N W Y Q Q K P G
91 |cag|tcc|att|tct|agc|tat|ctg|aat|tGG|TAC|Cag|caa|aag|ccg|ggT|
   |KpnI....
!
!      ----- FR2 ----->|--- CDR2 ----->|
!      42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
!      K A P K L L I Y A A S S L Q S
136 |aag|gct|ccg|aaa|ctg|tta|atc|tat|gcc|gct|tct|agt|ctg|cag|tct|
!
!      ----- FR3 ----->
!      57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
!      G V P S R F S G S G S G T D F
181 |ggT|gtt|ccg|TCT|AGA|ttc|tct|ggc|tct|ggT|tct|ggT|act|gat|ttt|
   |XbaI...
!
!      ----- FR3 ----->
!      72 73 74 75 76 77 78 79 80 81 82 83 84 85 86
!      T L T I S S L Q P E D F A T Y
226 |act|ctg|act|att|tcc|tct|ctg|caa|ccg|gag|gac|ttt|gct|acc|tat|
!
!      - FR3->|---- CDR3 ----->|--- FR4 ----->
!      87 88 89 90 91 92 93 94 95 96 97 98 99 100 101
!      Y C Q Q S Y S T P W T F G Q G
271 |tac|tgc|caa|cag|tct|tat|agt|act|ccg|tgg|act|ttc|ggT|caa|ggc|
!
!      ---- FR4 ----->|---- Ckappa----->
!      102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
!      T K V E I K R T V A A P S V F
316 |act|aaa|gtt|gag|att|aag|CGT|ACG|gtg|gct|gct|ccg|tct|gtc|ttc|
   |BsiWI..

```

Tabla 900: diversidad de CDR1

												Diversidad
Posición	24	25	26	27	28	29	30	31	32	33	34	
O12	R	A	S	Q	S	I	S	S	Y	L	N	
diversidad	2	2	1	1	3	1	2	2	4	1	3	576
permitido	Q	M			D		R	N	D		A	
					G				W		G	
									A			

Tabla 1000: Gran diversidad de CDR1

												Diversidad
Posición	24	25	26	27	28	29	30	31	32	33	34	
O12	R	A	S	Q	S	I	S	S	Y	L	N	
diversidad	3	2	4	1	5	1	4	5	5	1	6	72000
permitido	Q	M	E		D		R	N	D		A	
	E		R		G		E	E	W		G	
			Y		R		Y	R	A		D	
					Y			Y	R		R	
											Y	

Tabla 1100: diversidad de CDR2

POSICIÓN	50	51	52	53	54	55	56	Diversidad
O12	A	A	S	S	L	Q	S	
diversidad	2	1	1	3	1	2	2	24
permitido	D			N		E	T	
				T				

Tabla 1200: Gran diversidad de CDR2

POSICIÓN	50	51	52	53	54	55	56	Diversidad
O12	A	A	S	S	L	Q	S	
diversidad	4	1	4	6	1	4	5	1920
permitido	D		E	N		E	T	
	R		R	T		R	Y	
	Y		Y	E		Y	R	
				R			E	
				Y				

Tabla 1300: diversidad de CDR3

Posición	93	94	95	96	97	98	99	100	101	div.	tot.	
O12	Q	Q	S	Y	S	S	P	W	T			
diversidad	2	2	6	3	3	5	2	1	1	2160		
permitido	L	K	Y	D	N	T	S					
			H	N	Y	L						
			F			Y						
			A			F						
			D									

Tabla 1400: Gran diversidad de CDR3												
Posición	93	94	95	96	97	98	99	100	101	div.	tot.	
O12	Q	Q	S	Y	S	S	P	W	T			
diversidad	6	1	7	7	6	5	2	6	1	105840		
permitido	L		Y	D	N	T	S	F				
	E		H	N	Y	L		Y				
	R		F	R	D	Y		H				
	Y		A	A	R	F		L				
	A		D	L	A	E		I				
			R	S		R						

Ejemplo 7: Otros ejemplos de CDR3s de HC

Se examinó una colección de 22.063 Fabs con CDR3 distinta que se habían seleccionado de la biblioteca FAB-310 o FAB-410 y que fueron positivos para ELISA para al menos un antígeno. La utilización de las cadenas JH se muestra en la Tabla 1001; la parte FR4 de cada JH se muestra en cursiva. La Tabla 1010 muestra la utilización de aminoácidos en las CDR3s de HC. La Tabla 1020 muestra la distribución de longitudes de CDR3. La longitud de la mediana es 11,5.

La Tabla 1030 muestra la utilización de segmentos D en las CDR3s. Un segmento D identificado es 70% de los aminoácidos coincidentes; hay 5.654 casos (25,6%). Los Ds más usados fueron 3-3.2 (743, secuencia: YYDFWSGYYT), 3-22.2 (617, secuencia: YYDSSGYYY), 6-19.1 (441, secuencia: GYSSGWY), 6-13.1 (399, secuencia: GYSSSWY), y 4-17.2 (392, secuencia: DYGDY). De los Ds que contienen restos de Cys emparejados, 2-15.2 (secuencia: GYCSGGSCYS) fue la más usada; hubo 139 ejemplos, lo que representa 0,6% de la colección.

Cuando V o V::D se unen a J, hay a menudo edición del extremo 3' de V o V::D y el extremo 5' de J. La inspección de muchas secuencias de CDR3-FR4 muestra que a menudo hay una porción de JH que forma parte de CDR3. A menudo hay mutaciones en los restos de CDR3 que corresponden a los restos 1-9 de JH. Aquí, la porción de CDR3 que se piensa que deriva de JH se denomina el "muñón J". El JH usado en la cadena pesada se determina comparando cada uno de los restos de las seis cadenas de JH desde la posición 6 a 20 hasta la fusión de los últimos cuatro aminoácidos de CDR3 a FR4. Se selecciona el JH que tiene el número más pequeño de desemparejamientos. La secuencia de CDR3 se examina para un muñón J trabajando hacia atrás en el JH seleccionado desde la posición 9 hacia la primera posición del JH seleccionado comparando con CDR3 hasta que se termina la búsqueda por a) el extremo de JH, b) el extremo de CDR3, o c) dos desemparejamientos consecutivos. Si uno de los extremos de las cadenas y la última posición comparada es una coincidencia, entonces se incluye en

el muñón J. Si no, no lo es. La Tabla 1070 muestra varios ejemplos. La CDR está escrita anteriormente, el JH lo está debajo, y el muñón J está subrayado. En 1070 A, se comienza en 9, V coincide con V, y se continúa hasta la posición 6 con las coincidencias. La búsqueda se detiene en 4 debido al desemparejamiento doble. GMDV va a la pila del muñón J y GL entra en la pila de "Directriz". En 1070 B, la búsqueda termina con el extremo de JH6. Los restos subrayados pasan a la pila del muñón J, y EPIWG entra en la pila de Directriz. En 1070 E, la búsqueda termina debido al extremo de JH4, pero el resto final ensayado (D en la CDR frente a Y en JH4) es un desemparejamiento, y de esta manera el muñón J es FDS y DSGVVAAD pasa a la pila de Directriz.

La Tabla 1015 muestra la distribución de aminoácidos de CDR3s que no tienen segmentos D de los que se ha eliminado el muñón J. Obsérvese que la frecuencia de Tyr es mucho menor que cuando se compilan las CDR3s completas. Esto indica que Tyr entra en CDR3s en un mayor grado a través de la incorporación de segmentos D y muñones J. Estas Tyrs no están insertadas al azar, sino que ocurren en una secuencia que se ha seleccionado a lo largo de la evolución de mamíferos. Es una característica de la presente invención que se deberían insertar niveles elevados (más del 20%) de Tyr en las bibliotecas a través de la incorporación de Ds y muñones de J que contienen Tyr. En las posiciones directrices o de relleno DJ, se permite Tyr, pero no en más de 20%.

Tabla 1070: Ejemplos de asignación de muñón J

```

A)
  6      GLGMDV
JH6      YYYYYGMDVWGQGITTVTVSS
        123456789

B)
  13     EPIWGYYYYGMDV
JH6      YYYYYGMDVWGQGITTVTVSS

C)
  9      DFFTSYFDY
JH4      -----YFDYWGQGITLVTVSS

D)
  12     DRGVSLLGAFDI
JH3      -----AFDIWGQGITMVTIVSS

E)
  12     DSGVVAADFDS
JH4      -----YFDYWGQGITLVTVSS
        6789

```

La Tabla 1082 muestra la distribución del uso de aminoácidos en los muñones J de cada JH. Puesto que los JHs más comunes son JH3, JH4, y JH6, éstos son los JHs preferidos en los que construir bibliotecas. La Tabla 1082 muestra que la mayoría de los ejemplos de JH3 retienen la secuencia tetrapeptídica AFDI en CDR3. Con JH4, una mayoría retiene DY, y una gran fracción retiene la secuencia FDY en CDR3. Con JH6, una gran mayoría retiene la secuencia DV, una mayoría retiene la secuencia MDV, y una fracción sustancial retiene la secuencia GMDV. Una fracción no insignificante retiene la secuencia YGMDV, YYGMDV, o YYYGMDV.

Se incluyen en las bibliotecas de la presente invención las bibliotecas tales como 5.001 (Tabla 1097). La biblioteca 5.001 contiene LC y CDR1-2 de HC como se describe en otra parte en la presente solicitud. La biblioteca contiene un VH de HC (tal como 3-23) seguido de 6, 7, u 8 aminoácidos que permiten [GSRDLY] en la proporción mostrada en la Tabla 1097. En el muñón J, el aminoácido progenitor está presente 3, 4, 5, 6, 7, 8, 10 veces igual de probable que "otros" tipos de aminoácidos. Los "otros" tipos de aminoácidos comprenden Y, S, D, R, G. De este modo, en A6, se permite 7/12 A, más 1/12 de cada uno de Y, S, D, R, y G. En F7, se permite 7/12 F más 1/12 de cada uno de Y, S, D, R, y G. En D8, se permite 7/11 D más 1/1 de Y, S, R, y G. En 19, se permite 7/12 I más 1/12 Y, S, R, D, G. El aminoácido progenitor podría ser 5, 6, 7, 8, 10 veces más probable que los otros tipos de aminoácidos.

Se incluye en las bibliotecas de la presente invención la biblioteca 5.002 en la Tabla 1097. Esta biblioteca comprende CDR3 de longitud 13, 14, y 15, y ningún segmento D. Hay 6, 7, u 8 restos directrices que permiten G, S, R, D, L, o Y en las relaciones 1:0,57:0,46:0,42:0,36:0,35 o una aproximación razonable a ellas. La CDR3 se completa con una porción de JH6: YYYGMDV. El ADN que codifica la secuencia progenitora YYYGMDV se sintetiza con el aminoácido progenitor 5, 6, 7, 8, ó 10 veces más probable que los otros.

Se incluye en la biblioteca de la presente invención la biblioteca 5.003 en la Tabla 1097. FR3 es seguida de 4, 5, ó 6 restos directrices que permiten G, S, R, D, L, Y en la relación 1,0:0,57:0,46:0,42:0,36:0,35. A continuación sigue el segmento D 3-3.2; el ADN que codifica esta región favorece el aminoácido progenitor en 5 veces, y permite otros aminoácidos como Y, G, D, R, S. No hay relleno DJ, y los cuatro aminoácidos finales proceden del muñón J de JH3. El ADN que codifica el muñón J se sintetiza con el aminoácido progenitor 5 veces más probablemente que los otros: YSGRD.

La biblioteca 5.004 en la Tabla 1097 es una parte de la presente invención. Hay 2, 3, ó 4 restos directrices que permiten GSRDLY en las relaciones mostradas. El ADN que codifica la secuencia GYSSGWY se sintetiza de

manera que el aminoácido progenitor es 6-X tan probable como los otros, se permiten dos restos de relleno DJ con GSRDLY permitida en las relaciones 1,0:0,57:0,46:0,42:0,36:0,35. El ADN que codifica AFDI se sintetiza con el aminoácido progenitor 6-X tan probable como los otros.

La biblioteca 5.005 es parte de la presente invención. La biblioteca 5.005 comprende miembros de longitudes 11-14 de CDR3. Tras FR3, hay 0, 1, ó 2 restos directrices que permiten GSRDLY en las relaciones mostradas, seguido de ADN que codifica la secuencia progenitora GYSSGWY con variabilidad que permite YGSRD de manera que el aminoácido progenitor es 6-X tan probable como los otros tipos permitidos. Después de la región D, no hay ningún o hay un resto de relleno DJ que permite GSRDLY en las relaciones mostradas. Finalmente está JH3 con variabilidad en el muñón J (secuencia: YFDY) que permite YGSRD con el aminoácido progenitor 6-X tan probable como los otros tipos permitidos.

La biblioteca 5.006 en la Tabla 1097 es parte de la presente invención. La CDR3 puede tener una longitud 19-25. Hay cero a tres restos directrices que permiten GSRDLY en las relaciones mostradas. Tras la directriz está la región D 2-2.2. El ADN que codifica 2-2.2 se sintetiza de manera que el aminoácido progenitor es 6-X tan probable como los otros (es decir, YGSRD) excepto que los dos restos de Cys están fijos. Tras 2-2.2 hay cero a tres restos de relleno DJ que permiten GSRDLY en las relaciones mostradas. El ADN que codifica los primeros nueve restos de JH6 permite el aminoácido progenitor más YSGDR, siendo el tipo progenitor 6X más probable que los otros.

Tabla 1001: Utilización de JHs

			1111111112
JH	Número	%	12395678901239567890
JH1	1356	6,15	---AEYFQHWGQGTLVTVSS
JH2	1720	7,80	---YWYFDLWGRGTLVTVSS
JH3	5601	25,39	----AFDIWGQGTMTVTVSS
JH4	7658	34,71	----YFDYWGQGTTLVTVSS
JH5	1062	4,81	----NWFDPWGQGTTLVTVSS
JH6	4666	21,15	YYYYYGMDVWGQGTTLVTVSS
Total	22063		

Tabla 1010: Utilización de aminoácidos en CDR3 de HC

AA	Número	%	Rel up	Rel down
Y	42863	15,47	35,87	1,00
G	37512	13,54	31,39	0,88
D	34051	12,29	28,49	0,79
S	23068	8,33	19,30	0,54
F	17813	6,43	14,91	0,42
A	15150	5,47	12,68	0,35
R	14090	5,09	11,79	0,33
V	13834	4,99	11,58	0,32
L	12351	4,46	10,34	0,29
I	10014	3,61	8,38	0,23
P	9514	3,43	7,96	0,22
W	9340	3,37	7,82	0,22

T	7544	2,72	6,31	0,18
M	6093	2,20	5,10	0,14
E	6042	2,18	5,06	0,14
N	5901	2,13	4,94	0,14
H	4403	1,59	3,68	0,10
K	3147	1,14	2,63	0,07
Q	3097	1,12	2,59	0,07
C	1195	0,43	1,00	0,03
	277022			

Tabla 1015: Frecuencia de aminoácidos en Directriz de CDR3s que carecen de regions D				
AA	Número	Porcentaje	rel up	rel dn
G	23134	18,24	46,45	1,000
S	13555	10,69	27,22	0,586
R	10562	8,33	21,21	0,457
D	9704	7,65	19,49	0,419
L	8255	6,51	16,58	0,357
Y	8099	6,39	16,26	0,350
A	7188	5,67	14,43	0,311
V	6599	5,20	13,25	0,285
P	5768	4,55	11,58	0,249
W	4804	3,79	9,65	0,208
T	4769	3,76	9,58	0,206
E	4497	3,55	9,03	0,194
N	3733	2,94	7,50	0,161
F	3616	2,85	7,26	0,156
I	3464	2,73	6,96	0,150
H	2787	2,20	5,60	0,120
K	2460	1,94	4,94	0,106
Q	2124	1,67	4,27	0,092
M	1225	0,97	2,46	0,053
C	498	0,39	1,00	0,022
	126841			

Tabla 1020: Longitudes de CDR3s de HC

ES 2 528 963 T3

Longitud	Número	%
1	0	0,00
2	6	0,03
3	36	0,16
4	153	0,69
5	121	0,55
6	669	3,03
7	756	3,43
8	1066	4,83
9	2227	10,09
10	2701	12,24
11	2240	10,15
12	2071	9,39
13	2006	9,09
14	1594	7,22
15	1396	6,33
16	1254	5,68
17	1102	4,99
18	783	3,55
19	588	2,67
20	474	2,15
21	285	1,29
22	237	1,07
23	133	0,60
24	81	0,37
25	32	0,15
26	25	0,11
27	11	0,05
28	6	0,03
29	2	0,01
30	3	0,01
31	2	0,01
32	1	0,00
33	1	0,00
34	0	0,00

35	0	0,00
36	1	0,00
	22063	

Tabla 1030: Utilización de segmentos D.

Id	Número	Secuencia
1-1.1	29	GTTGT
1-1.2	6	VQLER
1-1.3	151	YNWND
1-7.1	34	GITGT
1-7.2	0	V*LEL
1-7.3	65	YNWNY
1-20.1	0	GITGT
1-20.2	0	V*LER
1-20.3	0	YNWND
1-26.1	48	GIVGAT
1-26.2	3	V*WELL
1-26.3	220	YSGSYY
2-2.1	0	RIL**YQLLY
2-2.2	102	GYCSSTSCYT
2-2.3	37	DIVVVPAAI
2-8.1	0	RILY@WCMLY
2-8.2	23	GYCTNGVCYT
2-8.3	1	DIVLMVYAI
2-15.1	0	RIL*WW*LLL
2-15.2	139	GYCSGGSCYS
2-15.3	12	DIVVVVAAT
2-21.1	0	SILWWSLLF
2-21.2	24	AYCGGDCYS
2-21.3	6	HIVVVTAI
3-3.1	28	VLRFLEWLLY
3-3.2	743	YYDFWSGYYT
3-3.3	15	ITIFGVII
3-9.1	41	VLRYFDWLL@
3-9.2	8	YYDILTGYYN
3-9.3	0	ITIF*LVII

ES 2 528 963 T3

3-10.1	26	VLLWFGELL@
3-10.2	136	YYYGSGSYYN
3-10.2	32	ITMVRGVII
3-16.1	0	VLSLRLGELSLY
3-16.2	109	YYDYVWGSYRYT
3-16.2	8	IMITFGGVIVI
3-22.1	0	VLL***WLLL
3-22.2	617	YYDSSGYYY
3-22.3	2	ITMIVVVIT
4-4.1	0	SLQ@L
4-4.2	75	DYSNY
4-4.3	165	TTVT
4-11.1	0	SLQ@L
4-11.2	0	DYSNY
4-11.3	0	TTVT
4-17.1	0	SLR@L
4-17.2	392	DYGDY
4-17.3	0	TTVT
4-23.1	0	SLRW@L
4-23.2	60	DYGGNS
4-23.3	16	TTVVT
5-5.1	25	VDTAMV
5-5.2	29	WIQLWL
5-5.3	292	GYSYGY
5-12.1	13	VDIVATI
5-12.2	0	WI*WLRL
5-12.3	200	GYSGYDY
5-18.1	0	VDTAMV
5-18.2	0	WIQLWL
5-18.3	0	GYSYGY
5-24.1	9	VEMATI
5-24.2	21	*RWLQL
5-24.3	44	RDGYNY
6-6.1	87	EYSSSS
6-6.2	122	SIAAR

ES 2 528 963 T3

6-6.3	1	V*QLV
6-13.1	399	GYSSSWY
6-13.2	170	GIAAAG
6-13.3	0	V*QQLV
6-19.1	441	GYSSGWY
6-19.2	104	GIAVAG
6-19.3	3	V*QWLV
7-27.1	257	LTG
7-27.2	0	@LG
7-27.3	64	NWG
ninguna	16409	

Tabla 1040: JH frente a longitud						
Longitud	JH1	JH2	JH3	JH4	JH5	JH6
1	0	0	0	0	0	0
2	1	4	0	1	0	0
3	20	2	3	9	0	2
4	75	3	10	45	8	12
5	47	6	10	38	8	12
6	273	14	43	280	26	33
7	88	27	194	337	30	80
8	134	43	243	503	41	102
9	121	70	855	886	61	234
10	116	693	623	979	68	222
11	105	81	675	1003	84	292
12	107	84	552	905	121	302
13	87	274	538	672	113	322
14	48	81	480	532	105	348
15	50	83	372	421	80	390
16	28	54	316	322	87	447
17	27	49	239	334	69	384
18	11	64	174	140	49	345
19	8	28	104	99	41	308
20	4	23	59	56	20	312
21	0	13	40	30	24	178
22	3	14	31	30	13	146

23	1	3	22	12	7	88
24	0	5	9	12	4	51
25	1	0	1	3	1	26
26	0	0	5	5	0	15
27	0	1	2	1	1	6
28	1	0	0	2	0	3
29	0	0	0	0	0	2
30	0	0	0	0	0	3
31	0	0	1	0	0	1
32	0	1	0	0	0	0
33	0	0	0	1	0	0
34	0	0	0	0	0	0
35	0	0	0	0	0	0
36	0	0	0	0	1	0

Tabla 1050: Utilización de aminoácidos en Directriz sin segmento D

AA	Número	%	Rel up	Rel dn
G	23134	18,24	46,45	1,00
S	13555	10,69	27,22	0,59
R	10562	8,33	21,21	0,46
D	9704	7,65	19,49	0,42
L	8255	6,51	16,58	0,36
Y	8099	6,39	16,26	0,35
A	7188	5,67	14,43	0,31
V	6599	5,20	13,25	0,29
P	5768	4,55	11,58	0,25
W	4804	3,79	9,65	0,21
T	4769	3,76	9,58	0,21
E	4497	3,55	9,03	0,19
N	3733	2,94	7,50	0,16
F	3616	2,85	7,26	0,16
I	3464	2,73	6,96	0,15
H	2787	2,20	5,60	0,12
K	2960	1,94	4,94	0,11
Q	2460	1,67	4,27	0,09
M	1225	0,97	2,46	0,05

ES 2 528 963 T3

C	498	0,39	1,00	0,02
	126841			

Tabla 1080: Dipéptidos en CDR3s de HC, parte 1							
YY	13565	FG	1073	PL	591	TV	397
FD	11637	RS	1072	TT	589	TP	390
DY	8337	SW	1014	ID	588	NA	389
SG	5979	DW	1003	DD	583	NS	388
GY	5805	LR	990	AS	570	ER	387
YG	5461	DG	989	KG	566	HG	386
DI	5448	PG	976	VD	556	VW	381
AF	4975	LL	974	VP	551	QL	378
DV	4968	AY	962	LT	540	RI	374
GG	4575	DR	923	LF	539	WN	365
SS	4491	VR	882	VL	539	YT	365
MD	4436	YM	877	FY	534	CS	360
GS	4047	AR	872	PD	533	DH	359
GM	3501	VV	869	RV	531	EA	359
YF	3438	YR	865	RF	525	WD	353
YD	3430	VA	857	AL	521	ES	350
RG	3118	RA	844	PS	510	FR	349
SY	2770	SP	820	EY	508	YC	343
GA	2611	GN	812	LW	508	PT	337
YS	2576	HY	809	PA	505	TL	326
DA	2285	SD	805	LP	500	KR	325
DS	2087	GI	804	VS	497	VF	324
WY	2079	NW	785	IR	493	MG	314
GD	2017	LS	760	YV	493	PN	313
GR	1985	LY	757	VY	478	RE	312
GL	1800	TY	749	IG	476	IV	311
DL	1777	PR	742	VT	475	KS	310
DF	1763	GE	737	TR	472	SC	310
GW	1725	SA	736	DN	471	FL	309
WS	1675	SF	728	SI	469	FF	306
AA	1671	PF	725	AD	462	CY	303

Tabla 1080: Dipéptidos en CDR3s de HC, parte 1							
LD	1651	ND	693	LA	459	SH	302
EG	1610	ST	684	PP	451	LK	300
AG	1606	GH	683	RT	451	IT	298
RY	1558	YP	676	DT	448	LE	298
DP	1547	WL	675	RW	447	FS	296
GV	1500	SN	667	GQ	446	ED	294
RR	1498	TS	652	QG	446	RK	294
LG	1387	RD	648	TD	446	HF	292
GF	1386	YA	648	TA	437	VI	290
VG	1366	SL	644	TF	426	RH	287
GP	1339	RP	643	GK	422	MV	285
WF	1282	YL	638	YW	421	KY	284
FW	1277	IA	634	HD	420	AI	282
NY	1271	RL	627	IL	417	HS	281
PY	1209	EL	622	LV	406	YH	281
GT	1194	YN	607	IS	402	LN	278
WG	1177	AV	605	NG	398	PV	276
SR	1162	AP	600	RN	398	QY	276
TG	1142	AT	592	SV	397	WA	271

Tabla 1080: Dipéptidos en CDR3s de HC, parte 2							
QH	267	KD	176	II	102	NQ	53
FQ	264	SK	176	HI	101	CF	51
LI	257	YK	176	KP	101	MP	50
EV	255	EF	174	MY	100	CP	49
AM	253	FN	174	RM	99	RC	47
DQ	250	HN	171	AQ	98	HE	46
HR	250	FH	165	EQ	96	VC	46
PH	248	YQ	165	QT	96	QI	45
AN	242	KN	164	LM	95	MN	44
WR	242	MA	163	HV	94	MF	43
NF	240	NN	160	IK	93	HQ	41
PI	239	KA	159	PM	93	CD	38
TN	239	SQ	157	QN	93	CL	38

Tabla 1080: Dipéptidos en CDR3s de HC, parte 2							
TI	238	PE	156	CG	91	NC	38
PW	229	WV	154	QF	91	HM	37
IP	228	EI	153	FI	90	FM	36
QR	227	TH	153	HW	90	ME	36
EW	225	FV	152	WH	90	MK	35
YI	221	AK	151	QV	89	QM	35
FE	220	TK	151	WI	89	NM	34
IY	220	WT	151	KH	88	KM	32
EP	219	PK	150	MI	88	TC	31
NR	217	KK	148	MS	87	CR	29
DM	214	IW	145	TQ	86	CV	25
FA	212	VH	145	NV	85	HC	25
AE	210	VE	141	EM	84	WM	25
IF	210	EE	138	HK	84	AC	24
QW	208	DE	136	IN	83	FC	24
YE	208	KL	136	NH	82	CA	23
FP	201	PQ	136	NI	82	CH	21
TM	201	QP	135	HT	81	CN	21
WE	201	SM	134	WK	79	MW	21
WP	201	QD	133	KF	77	PC	19
AH	199	QS	131	VM	73	LC	17
NP	198	VQ	130	MT	71	IC	16
VN	198	QQ	129	IH	69	MM	16
HA	196	WW	129	EH	68	MH	15
LH	196	NT	128	IE	67	WC	15
AW	193	DC	118	QK	65	EC	12
HP	192	KT	118	WQ	65	CK	10
HL	191	QA	118	GC	64	CW	10
RQ	191	NK	113	KE	61	MQ	10
TW	186	KW	112	KI	61	CI	9
EN	185	EK	109	CT	58	CC	8
LQ	182	FT	108	FK	58	CM	8
SE	180	KV	108	IM	57	CQ	6
VK	180	MR	105	KQ	57	QC	6

Tabla 1080: Dipéptidos en CDR3s de HC, parte 2							
ET	178	TE	104	ML	55	CE	5
DK	177	HH	103	QE	55	KC	5
NL	177	IQ	103	NE	53	MC	3

Tabla 1060a Longitudes de CDR3s de HC frente a qué segmentos D aparecen (si lo hacen) para longitudes 3-17

Nombre	Secuencia	Longitud														
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1-1.1	GTTGT	0	0	0	0	0	0	2	2	3	6	2	3	3	4	0
1-1.2	VQLER	0	0	0	0	0	0	0	0	1	0	1	0	1	2	1
1-1.3	YNWND	0	0	0	0	0	2	6	14	16	19	16	14	17	16	9
1-7.1	GITGT	0	0	0	0	0	0	0	1	2	7	6	4	4	0	4
1-7.3	YNWNY	0	0	0	0	1	0	2	5	7	8	8	6	5	9	4
1-26.1	GIVGAT	0	0	0	1	0	0	2	4	10	4	6	9	3	2	2
1-26.2	V*WELL	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1
1-26.3	YSGSYY	0	0	0	0	1	0	2	10	14	24	24	27	21	26	13
2-2.2	GYCSSTSCYT	0	0	0	0	0	0	0	0	0	2	2	9	15	15	11
2-2.3	DIVVPAAI	0	0	0	0	0	0	0	0	0	0	1	3	2	5	5
2-8.2	GYCTNGVCYT	0	0	0	0	0	0	0	0	0	1	0	2	4	3	4
2-8.3	DIVLMVYAI	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2-15.2	GYCSGGSCYS	0	0	0	0	0	0	0	0	1	3	5	12	10	25	22
2-15.3	DIVVAAAT	0	0	0	0	0	0	0	0	0	0	1	1	2	3	1
2-21.2	AYCGGDCYS	0	0	0	0	0	0	0	0	0	1	1	3	5	2	5
2-21.3	HIVAVTAI	0	0	0	0	0	0	0	0	0	1	2	0	2	0	1
3-3.1	VLRFEWLLY	0	0	0	0	0	0	0	0	0	0	1	2	2	2	3
3-3.2	YYDFWSGYT	0	0	0	0	0	0	0	1	5	8	22	38	44	72	69
3-3.3	ITIFGWII	0	0	0	0	0	0	0	0	0	0	0	1	0	2	5
3-9.1	VLRYFDWLL@	0	0	0	0	0	0	0	0	1	0	0	4	5	5	5
3-9.2	YYDILTGYN	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1

Tabla 1060a Longitudes de CDR3s de HC frente a qué segmentos D aparecen (si lo hacen) para longitudes 3-17

Nombre	Secuencia	Longitud														
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
3-10.1	VLLWFGELL@	0	0	0	0	0	0	0	0	2	1	2	4	2	3	5
3-10.2	YYYGSGSYN	0	0	0	0	0	0	0	2	4	7	10	13	15	18	14
3-10.2	ITMVRGVII	0	0	0	0	0	0	0	0	0	3	1	2	7	5	2
3-16.2	YYDYVWGSYRYT	0	0	0	0	0	0	0	0	0	0	0	1	7	7	7
3-16.2	IMITFGGVIVI	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2
3-22.2	YYDSSGYYY	0	0	0	0	0	0	0	0	6	30	45	56	59	108	101
3-22.3	ITMIVVVIT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4-4.2	DYSNY	0	0	0	0	0	3	3	4	6	14	4	7	4	10	6
4-4.3	TTVT	0	0	0	0	2	4	11	19	23	19	25	19	10	11	9
4-17.2	DYGDY	0	0	2	6	12	8	38	40	48	47	50	40	29	21	10
4-23.2	DYGGNS	0	0	0	0	0	2	7	4	5	17	4	8	5	1	1
4-23.3	TTVVT	0	0	0	0	0	0	2	0	1	1	2	1	2	0	0
5-5.1	VDTAMV	0	0	0	0	0	0	0	1	4	8	1	3	2	0	1
5-5.2	WQLWL	0	0	0	0	0	0	0	3	2	0	3	1	4	3	3
5-5.3	GYSYGY	0	0	0	0	1	6	9	20	43	29	27	22	32	26	27
5-12.1	VDIVATI	0	0	0	0	0	0	0	2	0	1	2	4	0	2	1
5-12.3	GYSGYDY	0	0	0	0	4	10	13	15	19	15	22	27	16	15	9
5-24.1	VEMAT I	0	0	0	0	0	0	1	1	0	0	6	0	0	0	0
5-24.2	*RWLQL	0	0	0	0	1	0	3	1	3	2	2	1	2	2	2
5-24.3	RDGYNY	0	0	0	0	0	0	0	1	8	12	6	7	3	2	1

Tabla 1060a Longitudes de CDR3s de HC frente a qué segmentos D aparecen (si lo hacen) para longitudes 3-17																										
Nombre	Secuencia	Longitud																								
6-6.1	EYSSSS	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17										
		0	0	0	0	0	0	0	9	7	16	19	13	2	4	2										
6-6.2	SIAAR	0	0	0	1	1	0	17	8	7	13	17	6	16	16	7										
6-6.3	V*QLV	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0										
6-13.1	GYSSSWY	0	0	0	0	0	1	6	11	35	40	56	44	50	42	35										
6-13.2	GIAAAG	0	0	0	0	1	2	18	14	15	20	20	15	16	14	11										
6-19.1	GYSSGWY	0	0	0	0	1	1	4	27	57	58	48	52	45	35	30										
6-19.2	GIAVAG	0	0	0	0	1	1	0	7	8	20	8	13	16	8	10										
6-19.3	V*QWLV	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0										
7-27.1	LTG	0	0	1	0	2	8	12	7	14	11	17	17	24	24	31										
7-27.3	NWG	0	0	0	1	2	11	6	5	10	6	7	5	7	1	0										
ninguno		36	153	118	660	726	1007	2063	2463	1851	1596	1502	1075	874	681	609										
Tabla 1060b Longitudes de CDR3s de HC frente a qué segmentos D aparecen (si lo hacen) para longitudes 18-32																										
Nombre	Secuencia	Longitud																								
1-1.1	GTTGT	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32										
		0	1	0	0	1	1	1	0	0	0	0	0	0	0	0										
1-1.2	VQLER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0										
1-1.3	YNWND	6	5	3	4	1	1	1	0	0	1	0	0	0	0	0										
1-7.1	GITGT	2	2	1	0	0	0	1	0	0	0	0	0	0	0	0										
1-7.3	YNWNY	5	2	1	1	0	0	0	0	0	0	0	1	0	0	0										
1-26.1	GIVGAT	0	0	3	0	1	0	0	1	0	0	0	0	0	0	0										

[illegible]

Tabla 1060b Longitudes de CDR3s de HC frente a qué segmentos D aparecen (si lo hacen) para longitudes 18-32

		Longitud															
3-9.2	YYDILTGYYN	3	2		0	0	0	0	0	0	0	0	0	0	0	0	0
3-10.1	VLLWFGELL@	3	4		0	0	0	0	0	0	0	0	0	0	0	0	0
3-10.2	YYYGSGSYYN	15	10		8	7	6	3	2	1	1	0	0	0	0	0	0
3-10.2	ITMVRGVII	3	3		6	0	0	0	0	0	0	0	0	0	0	0	0
3-16.2	YYDYVWGSYRYT	11	11		14	10	18	13	5	2	2	0	0	0	1	0	0
3-16.2	IMITFGGVIVI	1	3		1	0	0	0	0	0	0	0	0	0	0	0	0
3-22.2	YYDSSGYYY	77	54		28	22	18	8	1	2	1	1	0	0	0	0	0
3-22.3	ITMIVAVIT	1	1		0	0	0	0	0	0	0	0	0	0	0	0	0
4-4.2	DYSNY	7	2		2	1	1	0	1	0	0	0	0	0	0	0	0
4-4.3	TTVT	4	2		2	1	3	0	0	0	0	0	1	0	0	0	0
4-17.2	DYGDY	17	7		8	3	2	1	1	0	1	0	1	0	0	0	0
4-23.2	DYGGNS	3	1		0	2	0	0	0	0	0	0	0	0	0	0	0

Tabla 1060b Longitudes de CDR3s de HC frente a qué segmentos D aparecen (si lo hacen) para longitudes 18-32																									
		Longitud																							
4-23.3	TTVVT	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-5.1	VDTAMV	0	1	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-5.2	WIQLWL	3	3	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
5-5.3	GYSGY	13	18	7	2	2	6	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-12.1	VDIVATI	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-12.3	GYSGYDY	11	10	6	6	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-24.1	VEMATl	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-24.2	*RWLQL	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5-24.3	RDGYNY	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6-6.1	EYSSSS	9	3	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6-6.2	SIAAR	2	3	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6-6.3	V*QLV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6-13.1	GYSSSWY	24	22	17	4	5	1	3	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
6-13.2	GIAAAG	10	4	2	5	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6-19.1	GYSSGWY	23	25	13	7	7	3	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
6-19.2	GIAVAG	6	2	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6-19.3	V*QWLVL	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7-27.1	LTG	29	23	12	9	7	4	3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
7-27.3	NWVG	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ninguno		339	217	198	85	60	36	27	13	9	2	1	1	0	0	0	0	0	0	0	0	0	0	1	

Tabla 1082: Recuento de muñones J																								
JH1	---AEYFQHWGQGLTVTVSS 6,15%																							
	A		D		F	G		I		L	M											V		Y
		C		E			H		K			N	P	Q	R	S	T					W		
4	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	605
5	1	0	1	64	0	6	0	1	0	0	0	0	0	1	2	1	1	0	0	0	0	0	0	568
6	0	0	0	0	0	1	1	0	0	3	0	2	1	0	1	1	0	0	0	0	0	0	211	425
7	1	0	1	0	363	3	1	0	0	8	0	0	1	0	1	6	0	0	3	0	0	0	6	252
8	8	0	59	23	4	17	11	3	5	19	0	8	3	221	11	6	8	5	0	0	0	0	4	231
9	2	1	13	2	13	3	447	19	2	6	3	20	1	2	0	20	3	5	5	4	9	71		
JH2	---YWFYFDLWGRGTLTVTVSS 7,80%																							
	A		D		F	G		I		L	M											V		Y
		C		E			H		K			N	P	Q	R	S	T					W		-
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	159	1519
5	0	0	2	0	2	11	0	0	0	1	0	1	0	0	7	1	1	1	1	929	21	701		
6	3	2	9	0	40	5	7	4	1	9	0	7	1	0	1	11	2	0	0	1	1083	492		
7	1	6	1	0	1209	2	6	23	0	89	30	1	2	1	1	12	0	11	1	42	240			
8	31	2	1241	90	4	38	30	3	1	0	1	29	3	4	2	2	3	19	0	15	160			
9	3	1	9	3	34	2	26	17	5	1064	36	17	30	38	33	20	8	83	1	177	71			
JH3	-----AFDIWGQGTMTVTVSS 25,4%																							

Tabla 1082: Recuento de muñones J																															
	A		D		F	G		H		I		K	L	M		N		P		Q		R		S		T	V		Y	-	
			C		E							K													T			W			
6	4374	0	0	0	0	0		0		0		0	0	0		0		0	0	0	0	0	0	0	0	0	0	0	1074		
7	1	4	0	1	4612	2		0		8		0	56	3		2		1	1	1	0	29	2	15	12	74	625				
8	23	0	4765	51	0	28		14		0		3	0	0		15		2	1	1	2	2	6	0	4	531					
9	7	5	5	0	73	2		1		4439		4	64	64		43		2	1	11	54	49	113	2	18	491					
JH4	-----YFDYWGGGTLTVSS 34,7%																														
	A		D		F	G				I			L	M													V		Y		
			C		E			H				K				N		P	Q	R	S		T			W			-		
6	0	0	0	0	0	0		0		0		0	0	0		0		0	0	0	0	0	0	0	0	0	0	1975	5683		
7	9	1	10	7	3950	31		6		26		0	109	4		5		24	4	5	35	7	28	16	59	3322					
8	26	0	5991	32	5	91		19		2		1	6	1		21		7	0	8	13	3	6	2	14	1410					
9	5	18	17	0	119	2		0		0		14	0	2		64		15	16	10	216	11	3	6	6317	823					
JH5	----NWFDPWGGGTLTVSS 4,8%																														
	A		D		E	F	G			I			L														V		Y	-	
			C					H				K		M		N		P	Q	R	S		T			W					
5	0	0	0	0	0	0		0		0		0	0	0		274		0	0	0	0	0	0	0	0	0	0	764			
6	2	1	0	0	0	1		0		0		0	0	0		0		0	0	0	0	0	0	0	644	0	390				
7	1	1	0	0	768	2		1		1		0	13	0		0		0	0	0	1	0	2	0	4	244					
8	2	0	810	3	0	1		0		0		0	0	0		1		0	0	0	0	1	1	0	1	218					
9	3	0	3	0	3	0		3		0		0	4	0		4		814	1	0	14	2	0	0	187						
JH6	YYYYYGMIDWGQGTTVSS 21,1%																														

Tabla 1082: Recuento de muñones J																															
	A		D		F	G		H		I		L	M		N		P		Q		R		S		T		V		Y	-	
			C		E							K																			
1	0	0	0	0	0	0		0		0		0	0		0		0		0		0		0		0		0		690	3967	
2	4	0	15	0	16	31		12		2		1	4	0	8		7		4		14		18		7		4		1	1694	2815
3	3	0	19	3	14	23		16		1		5	9	0	16		9		5		12		20		4		1		4	2440	2053
4	4	0	14	2	31	16		25		3		2	10	0	13		6		5		15		35		5		5		4	2815	1647
5	2	1	9	1	23	19		21		1		1	3	1	15		3		1		7		26		3		0		1	3350	1169
6	69	5	14	1	4	3057		8		1		0	0	0	8		7		5		4		15		3		7		0	657	792
7	3	0	1	0	4	5		0		18		2	108	3866	0		2		2		3		1		7		18		3	1	613
8	7	0	4064	5	1	17		4		1		0	2	0	11		2		0		0		3		2		3		2	6	527
9	9	0	1	0	6	1		1		19		0	7	2	1		1		0		1		4		0		4092		0	1	511

Tabla 1097: Bibliotecas de CDR3 de HC

Biblioteca	Longitud de CDR3	Directriz	Secuencia de región D	Secuencia de relleno DJ	Secuencia de muñón J	Secuencia de FR4
5.001	10, 11, 12	6, 7, o 8 X X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	ninguna	ninguna	AFDI (diversidad en texto)	WQQGTMVTVSS (JH4)
5.002	13, 14, 15	6,7,8 X X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	ninguna	ninguna	YYYGMDV (diversidad en texto)	WQQGTTVTVSS (jh6)
5.003	18, 19,20	4,5,6 X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	YYDFWSGYT (3-3,2)	ninguna	YFDY	WQQGTLVTVSS (JH3)
5.004	15, 16, 17	2,3,4 X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	GYSSGWY (6-19.1)	2X,X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	AFDI	WQQGTMVTVSS (JH4)
5.005	11-14	0, 1, 2 X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	GYSSGWY (6-19.1)	0, 1 X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	YFDY	WQQGTLVTVSS (JH3)
5.006	19-25	0,1,2,3 X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	GYCSGGSCYS (2-2,2) (restos de Cys constantes)	0, 1, 2, 3 X = (1,0G, 0,57S, 0,46R, 0,42D, 0,36L, 0,35Y)	YYYYYGMVDV (AAs progenitores otros)	WQQGTTVTVSS (jh6)

REIVINDICACIONES

1. Una biblioteca de vectores o paquetes genéticos que presentan, presentan y expresan, o comprenden un miembro de una familia diversa de péptidos, polipéptidos y proteínas relacionados con anticuerpos humanos, y colectivamente presentan, presentan y expresan, o comprenden al menos una porción de la diversidad de la familia de anticuerpos, en la que los vectores o paquetes genéticos comprenden secuencias de ADN variegadas que codifican:

(i) una cadena pesada, que comprende:

(a) una CDR1 que tiene la siguiente secuencia de aminoácidos: $X_{31}YX_{33}MX_{35}$, en la que X_{31} , X_{33} y X_{35} son cualquier aminoácido excepto Cys o Met;

(b) una CDR2 que tiene la siguiente secuencia de aminoácidos: $X_{50}IX_{52}X_{52a}SGGX_{56}TX_{58}YADSVKG$, en la que X_{50} , X_{52} , X_{56} y X_{58} son cualquier aminoácido excepto Cys o Met, y X_{52a} se selecciona del grupo que consiste en Gly, Ser, Pro o Tyr; y

(c) una CDR3;

y

(ii) una cadena ligera.

2. La biblioteca de la reivindicación 1, en la que 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, ó 35 de las posiciones en la CDR3 de HC de la familia se varían entre los restos Tyr, Gly, Asp, Ser, y Arg, y en la que alrededor de 50% de todas las posiciones de la CDR3 de HC son Tyr, Gly, Asp, Ser, o Arg.

3. La biblioteca de la reivindicación 1, en la que 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, ó 35 de las posiciones en la CDR3 de HC de la familia se varían entre los restos Tyr, Gly, Asp, Ser, y Arg, y en la que un resto Arg está presente en una región de relleno entre V y D, entre D y J, o entre V y J.

4. La biblioteca de la reivindicación 1, en la que la CDR3 de cadena pesada (HC) consiste en 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, o 35 restos, cada uno de los cuales varía independientemente entre los restos Tyr, Asn, Asp, Ser, y Arg, en la que una región del marco (FR)3 está situada en el término amino de la CDR3, y una porción FR4 de una región JH está situada en el término carboxi de la CDR3; y en la que la FR3 es una FR3 de VH 3-23.

5. La biblioteca de la reivindicación 1, en la que más de alrededor de 20% de la secuencia de CDR3 de HC consiste en restos de Tyr, y en la que las regiones D y los muñones J de la CDR3 de HC consisten en más de alrededor de 20% de restos de Tyr.

6. La biblioteca de la reivindicación 1, en la que menos de alrededor de 20% de la secuencia de CDR3 de HC consiste en restos de Tyr, y en la que en las posiciones directrices o de relleno DJ de la CDR3 de HC consisten en menos de alrededor de 20% de restos de Tyr.

7. La biblioteca de la reivindicación 1, en la que la CDR3 de HC tiene tres aminoácidos de longitud y: el primer resto de la CDR3 de HC varía entre F, S, Y, D, y R en una relación 3:1:1:1:1; el segundo resto de la CDR3 de HC varía entre Q, E, R, S, Y, y L en una relación 3:1:1:1:1:1; y el tercer resto de la CDR3 de HC varía entre H, D, R, S, Y, y L en una relación 3:1:1:1:1:1, y en la que una FR3 está situada en el término amino de la CDR3, y el último resto de la FR3 varía entre K y R en una relación 3:1.

8. La biblioteca de la reivindicación 1, en la que la CDR3 de HC tiene una longitud de cuatro aminoácidos y: el primer resto de la CDR3 de HC varía entre L, S, Y, D, y R en una relación 4:1:1:1:1; el segundo resto de la CDR3 de HC varía entre L, S, Y, D, y R en una relación 4:1:1:1:1; el tercer resto de la CDR3 de HC varía entre W, S, Y, D, y R en una relación 4:1:1:1:1; y el cuarto resto de la CDR3 de HC varía entre F, S, Y, D, y R en una relación 4:1:1:1:1, y en la que una FR3 está situada en el término amino de la CDR3, y el último resto de la FR3 varía entre K y R en una relación 4:1.

9. La biblioteca de la reivindicación 1, en la que la CDR3 de HC tiene una longitud de dieciséis aminoácidos y: el primer resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el segundo resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el tercer resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el cuarto resto de la CDR3 de HC varía entre D, Y, S, R, y L en una relación 3:1:1:1:1; el quinto resto de la CDR3 de HC varía entre S, Y, R, D, y L en una relación 3:1:1:1:1; el sexto resto de la CDR3 de HC varía entre S, Y, R, D, y L en una relación 3:1:1:1:1; el séptimo resto de la CDR3 de HC varía entre G, A, S, Y, R, D, y L en una relación 3:1:1:1:1:1:1; el octavo resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el noveno resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el décimo resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el undécimo resto de la CDR3 de HC varía entre

A, S, Y, R, y D en una relación 3:1:1:1:1; el duodécimo resto de la CDR3 de HC varía entre E, R, S, Y, y L en una relación 3:1:1:1:1; el decimotercer resto de la CDR3 de HC varía entre Y, S, R, D, y L en una relación 3:1:1:1:1; el decimocuarto resto de la CDR3 de HC varía entre F, Y, S, R, y D en una relación 3:1:1:1:1; el decimoquinto resto de la CDR3 de HC varía entre Q, E, R, S, e Y en una relación 3:1:1:1:1; y el decimosexto resto de la CDR3 de HC varía entre H, E, R, S, Y, y L en una relación 3:1:1:1:1, y en la que una FR3 está situada en el término amino de la CDR3, y el último resto de la FR3 varía entre K y R en una relación 3:1.

10. La biblioteca de la reivindicación 1, en la que la CDR3 de HC tiene una longitud de dieciséis aminoácidos y: el primer resto de la CDR3 de HC varía entre G, S, Y, D, R, y L en una relación 3:1:1:1:1:1; el segundo resto de la CDR3 de HC varía entre Y, S, D, R, y L en una relación 3:1:1:1:1; el tercer resto de la CDR3 de HC es C; el cuarto resto de la CDR3 de HC varía entre S, Y, R, D, y L en una relación 3:1:1:1:1; el quinto resto de la CDR3 de HC varía entre S, Y, R, D, y L en una relación 3:1:1:1:1; el sexto resto de la CDR3 de HC varía entre T, Y, R, D, y L en una relación 3:1:1:1:1; el séptimo resto de la CDR3 de HC varía entre S, Y, R, D, y L en una relación 3:1:1:1:1; el octavo resto de la CDR3 de HC es C; el noveno resto de la CDR3 de HC varía entre Y, S, D, R, y L en una relación 3:1:1:1:1; el décimo resto de la CDR3 de HC varía entre T, Y, R, D, y L en una relación 3:1:1:1:1; el undécimo resto de la CDR3 de HC varía entre A, S, Y, D, R, y L en una relación 3:1:1:1:1:1; el duodécimo resto de la CDR3 de HC varía entre E, R, S, Y, y L en una relación 3:1:1:1:1; el decimotercer resto de la CDR3 de HC varía entre Y, S, D, R, y L en una relación 3:1:1:1:1; el decimocuarto resto de la CDR3 de HC varía entre F, Y, S, R, D, y L en una relación 3:1:1:1:1:1; el decimoquinto resto de la CDR3 de HC varía entre Q, E, R, S, Y, y L en una relación 3:1:1:1:1:1; y el decimosexto resto de la CDR3 de HC varía entre H, D, R, S, Y, y L en una relación 3:1:1:1:1, y en la que una FR3 está situada en el término amino de la CDR3, y el último resto de la FR3 varía entre K y R en una relación 3:1.

11. La biblioteca de la reivindicación 1, en la que la cadena ligera es una cadena ligera (LC) A27,

en la que la CDR1 de LC está diversificada, en la que la posición 27 es alrededor de 55% de Q, y alrededor de 9% de cada uno de E, R, Y, S, y L; la posición 28 es alrededor de 46% de S y alrededor de 9% de cada uno de N, T, Y, E, R, y L; la posición 30 es alrededor de 55% de S y alrededor de 9% de cada uno de D, N, R, T, e Y; la posición 30a es alrededor de 46% de S y alrededor de 9% de cada uno de G, N, R, T, Y, y D, y alrededor de 8% con ningún aminoácido; la posición 31 es alrededor de 44% de S y alrededor de 8% de cada uno de D, F, G, N, R, T, e Y; la posición 32 es alrededor de 44% de Y y alrededor de 7% de cada uno de F, D, L, N, Q, R, S, e Y; y la posición 34 es alrededor de 70% de A y alrededor de 15% de cada uno de S e Y, y

en la que CDR2 de la LC está diversificada, en la que la posición 50 es alrededor de 55% de G, y alrededor de 9% de cada uno de D, R, S, Y, y L; la posición 53 es alrededor de 52% de S y alrededor de 8% de cada uno de N, T, S, Y, E, y R; y la posición 56 es alrededor de 64% de T y alrededor de 9% de cada uno de E, R, S, e Y.

12. La biblioteca de la reivindicación 1, en la que la cadena ligera es una cadena ligera (LC) A27,

en la que la CDR1 de LC está diversificada, en la que la posición 27 es alrededor de 55% de Q, y alrededor de 9% de cada uno de E, R, Y, S, y L; la posición 28 es alrededor de 46% de S y alrededor de 9% de cada uno de N, T, Y, E, R, y L; la posición 30 es alrededor de 55% de S y alrededor de 9% de cada uno de D, N, R, T, e Y; la posición 30a es alrededor de 46% de S y alrededor de 9% de cada uno de G, N, R, T, Y, y D, y alrededor de 8% con ningún aminoácido; la posición 31 es alrededor de 44% de S y alrededor de 8% de cada uno de D, F, G, N, R, T, e Y; la posición 32 es alrededor de 44% de Y y alrededor de 7% de cada uno de F, D, L, N, Q, R, S, e Y; y la posición 34 es alrededor de 70% de A y alrededor de 15% de cada uno de S e Y, y

en la que la CDR3 de LC está diversificada, en la que la posición 91 es alrededor de 64% de Y, y alrededor de 9% de cada uno de F, E, R, y S; la posición 92 es alrededor de 52% de G y alrededor de 8% de cada uno de A, D, R, S, T, e Y; la posición 93 es alrededor de 52% de S y alrededor de 8% de cada uno de D, F, N, R, T, e Y; la posición 94 es alrededor de 55% de S y alrededor de 9% de cada uno de W, E, R, Y, y S; la posición 95 es alrededor de 64% de P y alrededor de 9% de cada uno de E, R, Y, y S, y alrededor de 8% con ningún aminoácido; y la posición 96 es alrededor de 55% de L y alrededor de 9% de cada uno de E, R, P, Y, y S.

13. La biblioteca de la reivindicación 1, en la que la cadena ligera es una cadena ligera (LC) A27,

en la que la CDR2 de LC está diversificada, en la que la posición 50 es alrededor de 55% de G, y alrededor de 9% de cada uno de D, R, S, Y, y L; la posición 53 es alrededor de 52% de S y alrededor de 8% de cada uno de N, T, S, Y, E, y R; y la posición 56 es alrededor de 64% de T y alrededor de 9% de cada uno de E, R, S, e Y, y

en la que la CDR3 de LC está diversificada, en la que la posición 91 es alrededor de 64% de Y, y alrededor de 9% de cada uno de F, E, R, y S; la posición 92 es alrededor de 52% de G y alrededor de 8% de cada uno de A, D, R, S, T, e Y; la posición 93 es alrededor de 52% de S y alrededor de 8% de cada uno de D, F, N, R, T, e Y; la posición 94 es alrededor de 55% de S y alrededor de 9% de cada uno de W, E, R, Y, y S; la posición 95 es alrededor de 64% de P y alrededor de 9% de cada uno de E, R, Y, y S, y alrededor de 8% con ningún aminoácido; y la posición 96 es alrededor de 55% de L y alrededor de 9% de cada uno de E, R, P, Y, y S.

14. La biblioteca de cualquiera de las reivindicaciones 1-10, en la que la cadena ligera tiene diversidad en una o más de CDR1 de LC, CDR2 de LC, y CDR3 de LC.

15. La biblioteca de la reivindicación 14, en la que la cadena ligera comprende una CDR1 de LC, una CDR2 de LC, o una CDR3 de LC de una LC de A27, en la que la CDR1 de LC está diversificada, en la que la posición 27 es alrededor de 55% de Q, y alrededor de 9% de cada uno de E, R, Y, S, y L; la posición 28 es alrededor de 46% de S y alrededor de 9% de cada uno de N, T, Y, E, R, y L; la posición 30 es alrededor de 55% de S y alrededor de 9% de cada uno de D, N, R, T, e Y; la posición 30a es alrededor de 46% de S y alrededor de 9% de cada uno de G, N, R, T, Y, y D, y alrededor de 8% con ningún aminoácido; la posición 31 es alrededor de 44% de S y alrededor de 8% de cada uno de D, F, G, N, R, T, e Y; la posición 32 es alrededor de 44% de Y y alrededor de 7% de cada uno de F, D, L, N, Q, R, S, e Y; y la posición 34 es alrededor de 70% de A y alrededor de 15% de cada uno de S e Y; en la que la CDR2 de LC está diversificada, en la que la posición 50 es alrededor de 55% de G, y alrededor de 9% de cada uno de D, R, S, Y, y L; la posición 53 es alrededor de 52% de S y alrededor de 8% de cada uno de N, T, S, Y, E, y R; y la posición 56 es alrededor de 64% de T y alrededor de 9% de cada uno de E, R, S, e Y, o en la que la CDR3 de LC está diversificada, en la que la posición 91 es alrededor de 64% de Y, y alrededor de 9% de cada uno de F, E, R, y S; la posición 92 es alrededor de 52% de G y alrededor de 8% de cada uno de A, D, R, S, T, e Y; la posición 93 es alrededor de 52% de S y alrededor de 8% de cada uno de D, F, N, R, T, e Y; la posición 94 es alrededor de 55% de S y alrededor de 9% de cada uno de W, E, R, Y, y S; la posición 95 es alrededor de 64% de P y alrededor de 9% de cada uno de E, R, Y, y S, y alrededor de 8% con ningún aminoácido; y la posición 96 es alrededor de 55% de L y alrededor de 9% de cada uno de E, R, P, Y, y S, o sus combinaciones.

16. Un método para preparar la biblioteca de una cualquiera de las reivindicaciones 1 a 15, comprendiendo el método titubeo digital.